

OM of: US-09-303-518D-649 to: Issued_Patents_AA:* out_format : pdf
Date: Jun 30, 2002 6:57 AM

Command line parameters:

```
Search information block:
Query: US-09-303-518D-649
Query length: 4374
Database: Issued Patents.AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 185.890000
```

score_list:

Strd Orig	ZScore	EScore Len	Documentation

[illegible]

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-266-3847-47	229.50	267.50	8.6e-09	235.00
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-08-968-687A-10	223.00	260.02	2.5e-08	212
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-409-995-4	218.50	255.21	5.1e-08	1912
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-685-447-6	218.50	255.21	5.1e-08	1912
/cgn2_6/ptodata/1/1aa/6A_COMB	pep:US-08-409-995-2	214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/1aa/6A_COMB	pep:US-08-685-447-2	214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-377-1155-32	214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-08-913-942-2	214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-266-347-32	214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-074-658-15	213.00	245.22	1.4e-07	2439
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-413-814-79	204.00	241.02	5.0e-07	1213
/cgn2_6/ptodata/1/1aa/5B_COMB	pep:US-08-387-942C-3	199.50	233.44	1.1e-06	1403
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-266-347-28	199.00	235.43	1.1e-06	1104
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-139-637A-351	193.50	238.15	1.9e-06	44
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-336-447A-9	193.00	229.24	3.2e-06	941
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-266-347-30	192.00	227.20	2.9e-06	1004
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-336-447A-13	188.00	223.44	6.6e-06	873
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-049-793-2	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-156-212-6	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-304-626-6	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-316-301A-6	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/5B_COMB	pep:US-08-611-928-6	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-173-891-6	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-076-137-6	187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-08-452-655B-2	186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/1aa/6A_COMB	pep:US-08-470-360-3	186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/1aa/6A_COMB	pep:US-08-471-491-3	186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-08-456-662-3	186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-07-741-940-7	186.00	207.70	1.5e-05	2842
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-289-548A-7	186.00	207.70	1.5e-05	2842
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-452-654-7	186.00	207.70	1.5e-05	2842
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-452-655B-2	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-452-655B-7	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/6A_COMB	pep:US-08-450-582-2	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/6A_COMB	pep:US-08-450-582-7	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/5B_COMB	pep:US-08-821-555A-7	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/5B_COMB	pep:US-09-003-687A-7	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-316-605-7	186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-336-447A-5	185.50	219.88	1.0e-05	831
/cgn2_6/ptodata/1/1aa/6B_COMB	pep:US-09-336-447A-1	184.50	219.34	2.1e-05	832
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-07-741-940-2	184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-289-548A-2	184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/1aa/5A_COMB	pep:US-08-452-655A-2	184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/1aa/5B_COMB	pep:US-08-370-323-2	184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/1aa/					

1279 TCGGCTGAATTAATGAAACAACTTGGCAAGCGCGGGCTTCATATACAGCA 1322
1415 G1ythrSerAspAsnThrThrPrrLysG1yAlaG1yValSerValAlaG1 431
1329 AGACAATACCGGTAACCTTGGAAATGAACGCGCGTGGCAACAGCCCGCT 1378
431 uG1yLThrValThrTrpLysValHisnProGlnTrpAspArgLeuA 448
1379 CCAAAATCGGCAAAAGCGACGCTGCACGTTCAACGCCAAAGGGAAACCA 1428
448 lAluStlG1yLysG1yThrLleuAlaG1uG1yThG1yAspAsnLys 464
1429 GCGTCGATCAGCGCTGGCGCAGCGTACAGTCATTTGGATTCACACAGCA 1478
465 G1ySerLeuLysValG1yAspG1yThrVal11leuLysG1nG1nThrAs 481
1479 CGATTAAGGCAAAAAACAAACGCTTTAGTAAATCGGCTTGCTACCGGCA 1528
481 nG1ySerG1y...G1nHis1AlaPheAlaSerValG1y11leuAlaSerG1yA 497
1529 GGGGTACGGGCAACGATGGCGGATTAACAGTTCACCTGACCCGCAACATC 1578
497 rGerThrLleuVal11leuAsnAspAsLysG1nValAspProAsnSer1le 513
1579 TAATTCGGCTTTGCGCGCGAGCGTTGGATTAAACGGGCAATTCGCTTTC 1628
514 TyrPheG1yPheArgG1yLysTrgLeuAspLeuAsnG1yAsnSerLeuTh 530
1629 GTTTCACCGGATTCGAAATACCGATGAAGGGCGCATGATTTGTCAACACA 1678
530 rPheAspHis11leuAlaGsn11AspAspG1yLAlaTrgLeuValAsnHisA 547
1679 ATCAAGACAAAGAATACCGCTTACCATTTACAGCG... 1713
547 smethThrAsnAlaSerAsn11ethr11leThnG1yLysLeu11leThr 563
1713 1713
564 AspProAsnThr11leThrProTrgAsn11leAspAlaProAspG1uAspAs 580
1713 1713
580 nProTrgAlaPheArgArg11leLysAspG1yG1nLeuTrgLeuAsnL 597
1713 1713
597 euG1uAsnTrgThrTrgTrgAlaLeuArgLysG1yAlaSerThrArgSer 613
1713 1713
614 G1uLeuProLysAsnSerG1yG1uSerAsnG1uAsnTrpLeuTrgMetG1 630
1714 ...AATAAGATATTGGTACAAACCGCGCAATTAACAGCTGGATGACA 1759
630 LysTrnSerAspG1uAlaLysArgAsnValMetAsnHis11leAsnSng 647
1760 AAAAAGAAATTCCTACAAACGGTTGGCTTGGCGGAAGAAGATACGACCA 1809
647 LuTrgMetAsn11yPheAsnG1yTrpPheG1yG1uG1nG1u...G1yLys 662
1810 ACGAAGCGGCGCTCAACCTTGTTTACACAGCCCGCGCAGACAGCGCAC 1859
663 AsnAsnG1yAsnLeuAsnValThrPheLysG1yLysSerG1uG1nAsnAr 679
1860 CCTGCTGCTTCCGGCGGACAAATTTAAACGGCAACATCAGCAACAAACA 1909
679 gPheLeuLeuThrG1yG1yThrAsnLeuAsnG1yAspLeuThrValG1uL 696
1910 ACGGCAAACTGTTTACGGCGCAGACAAACACCGCGCTCAATTCAT 1959
696 yGcG1yThrLeuPheLeuSerG1yArgProThrProHisAlaArgAsp1le 712

1960	TTAAACGACCATGGTGGCAAAAAGAGGC.....ATTCCCTGGCGGGA	2007
713	AlaLylIleSerTlrLysAspProHisPheAlaGluAsnAsnL	729
2004	AATCGTGTGGACAGACGACTGGATTCACGCGACATTTAAAGCGGAAACT	2055
729	ValValValGluAspSprPrIleAsnPrIleAsnPrIleValIleThrIhm	746
2054	TTCCAAATTTAAAGCGGACAGCGCGTGGTTCC...CGCAATGTTGCCAA	2100
746	eLAsnValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsn	762
2101	GTCGAAGGCGATTGGCATTTGACCATTCACGGCCCAAGCAGTTTGGTGT	2150
763	IleIleHisSerHisIleThrAlaSerAsnLysAlaGluValHisIleGlyTyr	779
2151	CGCACCGCATCAAAAGCCACACCAATGTGTACAGCTTGGCGAGTGGAGGCTC	2200
779	r.....LysThrGlyAspThrValGlyValArgSerAspTyrThrGlyT	794
2201	TGACAAATTTCTGTGCGAAAAAACCATTCAGCGAGTAAGTAACTGTCTCTCA	2250
794	YrValThrCysThrThrAspLysIleSer...AspLysAlaLeuAsnSer	809
2251	TTGACATGAAGCGACATTCAGCGGCAATGTGATCTGTGCGATCAGCTCA	2300
810	PheAsnProThrAsnLeuThrArgGlyAsnValAsnLeuThrGluSerAlaAs	826
2301	TTTAAATCTCACAGGGCTTGGCCACACTCAACGGCGCATCTTACTGCAATG	2350
826	nPheValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArg	842
2351	GCGATACACGTTATTAACAGTCAGACCAACGGCCAAACCGGCAACCTT	2400
842	LysAsnSerGluValArgLeuThrGluAsn.....	851
2401	AGCCTGTGGCGCAATGCCACAGCACATTTAATCAAGCACATTAAACGG	2450
851	851
2451	CAACAGATCGGCTTGGGCAATGCTTCATTTAATTAAGCGACCGCG	2500
851	851
2501	TACAAAGCGGCACTGACGCTTCCGGCAGCGCTAAGCAACGTAAAGC	2550
851	851
2551	CATTCGCACTCAACGGTAATGTCTCCTAGCGGATTAAGGCAATTTCCA	2600
851	851
2601	TTTGAAGACGCCGCTTACCGGCAAAATCAGCGCGGCAAGATACGG	2650
851	851
2651	CATTACATTTAAAGACAGACGAGATGAGCGTCCGCTCAGCGACGAATTA	2700
852SerHisPrHisLeuThrGlyAsnSerAspVal	862
2701	GGCAATTTAAACCTTGACACAGCGCCACTTACCTCAATTCGGCTATTCG	2750
863	HisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAs	879
2751	CCACGATGGCGGACGAGGCGCAACCGCGCATGTCGACAGATGCGCGCGCC	2800
879	nSerAsn.....	881
2801	GCGCTTGGCGCGCTTGGCGGCTTCCATTATGCGGTTACACGCGCACT	2850
882AsnValIleThr.....	884
2851	TGCGTAGAATCCGTTTCAACAGCTGACGGTAAACGGCAATTGAACGG	2900

[illegible]

1177	SeArSVaIGluGInPrOvAlThrGluSeThrThrValAsnThrGluAs	1193
3421ACCGGCGCGGTACCAGCGCGCTTCC	3445
1193	nSeArVaLaIGuaInPrOJuaSnThrThrProAlaThrThrGln...P	1209
3446	CC.....CGCGCGCGCGCG	3459
1209	ThrValAsnSerGluSeArSInLysProLysAsnAlaGlnHSAATG	1225
3460	gccccgcggagatttccg...CACTGCACCCCAACCGGACGCCAAC	3506
1226	SeArVaLaIGuAInPrOJuaSnThrThrProAlaThrThrSeArS	1242
3507	GCAGGCC.....GACATGATACGGCTATGCSAATA	3538
1242	nAsPArSeThrThrValAlaLeuCyASpLeuThrSeThrAsnThrAsnA	1259
3539	GCGGTTTGAGTAATTTTCCGCC.....ACGCTCAACAGC	3573
1259	LaValLeuSeArSArPaLaGAlaLaGAlaInPheValAlaLeuAsnVal	1275
3574	gTTTTGGCGGTACAGAGCAATTAACGGCGTA...TTTGGCGGAAGCG	3620
1276	GIyLaVaLaVaLSeArGlnHSASeArGlnLeuGluLeuVaSnAsnGluG	1292
3621	CGCCAGCGCGGTTTGAGCAAGCGGATCGGGAGACACCAACATACGGTT	3670
1292	yGlnTyArSInValThrPValSeArSInThrSeArMeLAsnLysAsnTySeArS	1309
3671	CGCAAGATTTCCGGCCCTACCGCCCAACAACCGACTCGGCAATGGT	3720
1309	eSeArGlnTyArGaRgPheSeArSeTySeThrGlnThrGlnLeuGln	1325
3721	ATCGAGAAACCTCGGAGCGAG...CGCGTCGCGATCCGTTTTGCA	3767
1326	TPrSePAlnThrThrLeSeArSInSInValGlnLeuGlnGlyValAlaPheGlnTy	1342
3768	CAACCGGACGGAAACACCTTCGACGACGCGACTCGGCACTCGGACGCG	3817
1342	rVaLaRgSaSeArSInAsnPheAsPryLaLThrSeTySInThr...L	1358
3818	TTGGCCACGGCGCGCTTTCCGGCAATACGCAATCGC...AGGTTTAC	3864
1358	eVaLaGlnVaLaSnPheTySeTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy	1374
3865	ATCGGCATCGACCGCGCGCGGGGTTTTTACGACGCGGACCTTTCAGACGG	3914
1375	LeuTyThrLeAsPrenGlnTyGlyTyPheGlnSerTyLeuGlnThrAs	1391
3915	CATCGGAGGCAAAATCGCGCGCGCGCTGCGATTACGGCATTCAGCGAC	3964
1391	nHSAInAlaLysPheAlaRgAlaThrAlaGlnPheGlyLeuThrAlaG	1408
3965	GATACCGCGCGGTTTCGCGGATTTGCGGATCGAACCGCATTCGCGCA	4014
1408	LyLaSaLaPheAsnLeuGlnTyAsnPheGlyThrProIleValGlyAla	1424
4015	ACGGCGATTTCGTCACAAAGCGGATTACCGCTACGAAGACGTATAT	4064
1425	ArGlyTySeTy	1441
4065	CGGCAACCGCGCGCTTCATTCACCGCTACCGCGCGGATTAAGGACG	4114
1441	eLysVaLaSnProIleSeArTyAlaLysThrAlaPheAlaGlnValaLysAs	1458
4115	ATTATTCATTCACAAACCGCGCAACATTCGATCAGCGCTTATTGGAC	4164
1458	eTyTyThrTyThrTyHSAInGlu...PheSeTyAlaThrProIleLeuSeAr	1473
4165	CTGTCGATACCGATCGCGCTGGGCAAGTCCGAACCGCGCTGATAC	4214
1474	AlaArGlyTy...AsPAlaAsnGlnGlySeGlyTyTyTyTyTyTyTyTyTyTyTy	1489

851 851
2501 TCAAAAGCGAGTGTGACGTTTCCGGCAGCGTAAGCAACGTAAGC 2550
851 851
2551 CATTCGCAGCTACAGGTAATGTCTCCCTAGCCGATTAAGCAGTATTCGA 2600
851 851
2601 TTTGGAAGCAGCCGCTTACCGGACAATACGCGCGCAAGATACGG 2650
851 851
2651 CATTAACCTTAAGAAGACGGAATGAGCGTCGCGTACGCGCAGGAATTA 2700
852 862
2701 GGCAATTTAAACCTTGACAAAGCCACATTAACCTAATTCGCGCTATCG 2750
863 HtSGInLeuAspLeuAlaAsnGlyHtSInLeuAsnSerAlaAspAs 879
2751 CCACGATCGCGGAGGGCGCAACCGCAGTCGACAGATGCGCGCCG 2800
879 nSerAsn..... 881
2801 GCCGTTCCGCGGTTGCGCGGTTCCCTATTATCCGTTACACCGCCACT 2850
882 884
2851 TCGGTAGATCCCGTTGACACAGCGTGACGTAACGGCAATGTAAGC 2900
885 896
2901 TCAGGACATTCGCGCTTATGTGAGACTCTGCGCTACCGCAGCGCA 2950
896 yAsnGlySerPheTyrLeuThrAspLeuSerAsnGlySInGlyAspL 913
2951 AATTGAACTGCGGGAAGTTCCGAAGGCACTTACACTTGGCGGTCAAC 3000
913 ySValValAlThrLySerAlaThrGlyAsnPheThrLeuGlnValAla 929
3001 AATACGGCAGAACTCGCAAGCCTCGACAATTTGACGTAAGGAGG 3050
930 AspLySInGlyGluPro...AsnHisAsnGluLeuThrLeuPheAspAl 945
3051 AAAAGACAAACACCGCTGTCGGAACCTTAATTTCACTCCCTGCAAAAGC 3100
945 a.....SerLySAlaGlnArgAspHisLeuAsnValSerLeuValGlyA 960
3101 AACAGTCGATGCGCGCGCGGTACCAACTCATCCGCAAAAGCGC 3150
960 snThrValAspLeuGlyAlaThrLySInGlyLeuArgAsnValAsnGly 976
3151 GAGTTCGCGCTGCAATTCGCGTCAAGAACA..... 3183
977 ArgTyAspLeuTyrAsnProGluValGluLySArgAsnGlnThrValAs 993
3183 3183
993 pThrThrAsnThrThrProAsnAsnIleGlnAlaAspValProServ 1010
3184GAGCTTCGACAACTCGGCAAG..... 3207
1010 aIProSerAsnAsnGluGlnIleAlaArgValAspGluAlaProValPro 1026
3208GAGAGAG 3215
1027 ProProAlaProAlaThrProSerGluThrThGluThrValAlaGluAs 1043
3216 CAATAAAAGCGGGAATAA.....GACAAAGCGCAAAAGCTTGAGC 3256
.....
1043 nSerLySInGluLeuSerLySThrValGluLySAsnGluGlnAspAlaThrG 1060
3257 CGCTGATTGCGCGCGCGCGATCCGTCGA..... 3288
1060 lThrThrAlaGlnAsnArgGluValAlaLySAlaLySAsnVal 1076
3289 AAGACAGAAAGCTTGCGCAACCGCGCGGAGCAGCGGGAATAATC 3338
1077 LySAlaSnThrGlnThrAsnGluValAlaGlnSerLySInGlyThr 1093
3339 CGGCATTATGACGCGGAG.....GAAG 3361
1093 sGluThrGluThrThrGluThrLySgluThrAlaThrValGluLySglu 1110
3362 AGAAAAAGCGGTGACGCGGATTAAGACACCGCTTGCGGAAA..... 3405
1110 lutySAlaLySValGluThrGluThrGlnGluValProLySValThr 1126
3406CAGCGCAGAC 3416
1127 SerGlnValSerProLySInGluGlnSerGluThrValGlnProGlnAl 1143
3417 GGA..... 3420
1143 aGluProAlaArgGluAsnAspProThrValAsnIleLySgluProGlns 1160
3420 3420
1160 ergInThrAsnThrThrAlaAspThrGluGlnProAlaLySgluThrSer 1176
3420 3420
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3421ACCGCGCGGTACCGCCCTTC 3445
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
3446 CC.....CGCGCGCGCGC 3459
1209 roThrValAlaSerGluSerSerAsnLySProLySAsnArgHisAlaGAG 1225
3460 GCCCGCGGATTGCGG...CACTGCAACCCCAACCGCAGCCCAAC 3506
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3507 GCAGCGC.....GACCTGATCAGCCGTTATGCCAATA 3538
1242 naSPArgSerThrValAlaLeuCySAspLeuThrSerThrAsnThrAsna 1259
3539 GCGGTTGAGTGAATTTCCGC.....ACGTCACACAC 3573
1259 lValIleGlnSerAspAlaArgAlaLySAlaGlnPheValAlaLeuAsnVal 1275
3574 GTTTTCGCGCTACAGACGCGCATCCGGGACACCAACTACCTACCT 3620
1276 GlyLyAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluG 1292
3621 CCGCAACGCGCTTTGACAAAGCGCATCCGGGACACCAACTACCTACCT 3670
1292 yGInTyrAsnValTTPValSerAsnThrSerMetAsnLySAsnTyrSerS 1309
3671 CGCAAGATTTCCGCGCGCTACCGCCCAACAACGACCTGCGCAAAACGCT 3720
1309 eSerGlnTyrArgArgPheSerSerLySInGlnThrGlnLeuGly 1325
3721 ATGCAGAAAAACCTGCGCAGCGG...CGCGTGCAGTCCTGTTTGCA 3767
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3768 CAACGCGACGAAACACCTTGACGAGGCGATCGGCAACTGCGCAGCGC 3817
1342 rValArgAsnSerAsnAsnPheAspLySAlaThrSerLySAsnThr...L 1358

```

3818 TTGCCCCAGCGCCGCTTTGGGCAATAGCGCATGCAC...AGTTCTAC 3864
      |||||||
1358 euAlaGlnValAsnPhetYSerLysTyrAlaAspAsnHisTrrPyl 1374
      :|||
3865 ATGGCATTCAGCGCGCGGCTTTTACAGCGCGAGCTTTCAGACGG 3914
      :|||
1375 LeuGlyLeuSpleuGlyTyrGlyLysPheGlnSerLysLeuGlnHis 1391
      :|||
3915 CATGGAGGCAAAATCCGCGCGCTGCTGCATTCAGCGCATTCAGCGAC 3964
      :|||
1391 nHISAsnAlaLysPheAlaArgHisTrrAlaGlnPheGlyLeuThrAlaG 1408
      :|||
3965 GATACCGCGCGGCTTCGCGGATTCGCGCATCGACCGGACATCGCGCA 4014
      :|||
1408 LyLysAlaPheAsnLeuGlyAsnPhetYLeuThrProLleValGlyVal 1424
      :|||
4015 ACGGCGTATTCGTCGCAAAAAGCGATTCACGCTACGAAAACGTCATAT 4064
      :|||
1425 ArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArg 1441
      :|||
4065 CGCCACCGCGCGCTTCGATTCACACGCTACGCGCGGCGATTAAGCGAC 4114
      :|||
1441 eLysValAsnProLleSerValLysThrAlaPheAlaGlnValAspLeuS 1458
      :|||
4115 ATTTTCATTCACAAACGCGCGCAACATTCATTCACGCTTATTTGAGC 4164
      :|||
1458 eTrrYrrTrYrrHisLeuGlyGlu...PheSerValThrProLleLeuSer 1473
      :|||
4165 CTGTCTATACCGATGCGCTTCGCGGCAAAAGTCCGAAACGCGCTCATAC 4214
      :|||
1474 AlaArgTyr...AspAlaAsnGlnGlySerGlyLysLleAsnValAsnG 1489
      :|||
4215 CGCGGTATTCGCTCAGATTCGCGCAAAACGCGAGTCGGAATGGGCGC 4264
      :|||
1489 YTrAspPheAlaLysAsnValGlnAsnGlnGlnGlnGlnTrAsnAlaGly 1506
      :|||
4265 TAAACGCGCGAATCAAGGTTTCACGCTGCTCCACGCTCGCGCGCGC 4314
      :|||
1506 euLysLeuLysTyrHisAsnValLysLeuSerLeuLleGlyGlyLeuThr 1522
      :|||
4315 AAAAGCGCGCGCAACGCGGCAACGCGGCGCGCATCAATTAAGGCTA 4364
      :|||
1523 LysAlaLysGlnAlaGlnLysGlnLysThrAlaGlnLeuLysLeuSer 1539
      :|||
4365 C 4365
      :|||
1539 e 1539

```

```

seq_name: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:US-08-296-791-4
seq_documentation_block:
; Sequence 4, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehm, Hombach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RET/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-4

```

```

alignment_scores:
Quality: 1579.50 Length: 1735
Ratio: 1.704 Gaps: 53
Percent Similarity: 53.429 Percent Identity: 26.686

```

```

alignment_block:
US-09-303-518D-649 x US-08-296-791-4 ..

```

```

Align seg 1/1 to: US-08-296-791-4 from: 1 to: 1545

```

```

64 CGCTTCGCGCGCTTACTTACGATATGCTGCTGCGCATTTCC 113
      :|||
5 LysPheLysLeuAsnPhetLleAlaLeuThrValAlaTrrAlaLeuThr 21
      :|||
114 CCAACGCTGCGCGGAGCACACTTATTCGCGCATCACTCAATCAATATATC 163
      :|||
21 OTrThrGlnAlaAlaLeuValArgAspValAspValAspTrGlnLlePhe 38
      :|||
164 GCGACTTTCGCGCAAAATTAAGGCAAGTTTGCAGTCGCGGCGCAAGATAT 213
      :|||
38 rGAspPheAlaGlnAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
      :|||
214 GAGGTTTACAAACAAAAGGAGTGTGTCGCGCAATCAATGACAAAAGC 263
      :|||
55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnG 71
      :|||
264 C...CGATGATTAATTTCTGTGTGTGTC...CGTAAAGCGCGGCGG 307
      :|||
71 YLleProMetLleAspPheSerValValAspValAspLysArgLleAla 88
      :|||
308 CATGTGCGCGCATCAATATATTCGAGCGTGCGACAT..... 345
      :|||
88 hTrLeuValAsnProGlnTrrValValGlyValLysHisValSerAsnGly 104
      :|||
346 .....AAGCGCGCTTAACACGTTGA 368
      :|||
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
      :|||
369 TTTTGTGCGGGAAGAAATCCGATCAACATGCTTTACTTTATAAA 418
      :|||
121 nAlaLysAlaHis...ArgAspValSerSerGlnLysAsnArgTrrTrr 137
      :|||
419 TTGTGAACGGAATATATAT.....AAGCAGGAGCT 450
      :|||
137 hValGlnLysAsnGlnTrrProThrLysLeuAsnGlyLysAlaValThr 153
      :|||
451 .....AAGCGCATCTTATGCGCGGATTAATCATATGCGCGTTT 491
      :|||
154 ThrGlnAspGlnAlaGlnLysArgGlnAspTrrTrrMetProArgGle 170
      :|||
492 GCATAATTTTGCACAGATGCGAAGCTGTGA...ATGACCATTTTAA 538
      :|||
170 uAspLysPheValThrGlnValAlaProLleGlnLleAsnThrAspSers 187

```

539 TGGATGGCGGAAATATATGATCAAAATATATACCTGACCGGTGTGT 588
187 eThrAlaGlyThrIAsnAsnLysAspLysTrpProLysPheValArg 203
589 ATTGGGCGGCGAGCAATAT..... 609
204 LeuGlySerGlyThrGlnPheIleThrGlnAsnGlyThrArgTyrGluLe 220
610 .TGGGATCTGATGAAGATGAGCCCAATACCGCAAGATTCAATCAT 657
220 uThrLeuGlyLysGlyGlnLysSerAsp...AlaGlyGlyTyrAsnL 236
658ATTGCAAGTGGATTTCTTGGCTTCGTGGTGGCAATACCTTT 699
236 euLysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrProTyr 252
700 GCACAAATGATCAGTGGTGGCACAGTCAACTAGTAGT..... 741
253 GluValAsnHisGluAsnAspLysLeuIleGlyPheGlyAsnSerAsnAs 269
742GAAAAATTAAACATAGCCCATATGGTTT 772
269 nGluTyrIleAsnProLysGluIleLeuSerLysPro.....L 283
773 TACCAACAGGAGGCTCATTTGGCGACAGTGCCTCACCAATGTTATCTAT 822
283 euThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyr 299
823 GATGCCCAAAAGCAAAAGTGGTATTAATGAGGTATGCAACGGCGCA 872
300 AsparGluLysGlyLysTrpLeuPheLeuGlySerTyrAspTyrTPAI 316
873 CCCCTATATAGCAAAAGCAATGGCTTCAGCTGGTTGTAAGATTGGT 922
316 aGlyTyr.....AsnLysSerTrpG 324
923 TCTATGATGAATCTTTCGTGAGATACCATTCAGTATTTCTACGACCA 972
324 InGluTyrAsnIleTyrLysProGluPheAlaGlyLysIleTyrGln 340
973 CGTCAAAATGGG.....AATACCTTTTACGACGATAA 1007
341 TyrSerAlaGlySerLeuIleGlySerLysThrAspTyrSerTrpSer 357
1008 TAATGGCACAGAAATCAATGCCAATACATGAACACATTCCTGCTTA 1057
357 rAsnGly..... 359
1058 ATAGATTAAACACAGAACCTT.....CAATGTTTAAATGT 1095
360LysThrSerThrIleThrGlyGlyLysSerLeuAsnVal 373
1096 TCTTATCCGAGACAGCAAGACCTGTTATCATGCTGCAGGTGGTGT 1145
374 AspLeuAlaAspGlyLysAspLysPro..... 382
1146 CAACAGTATAGACCCAGACTGAATTAATGCAAAATATTTCTTATATG 1195
383AsnHisGlyLysSerValThrPhe.... 390
1196 AGAAGAAAGGCAATGATGATCTACCGACACATCAATCAAGTGT 1245
391 ..GluGlySerGlyThrLeuThrLeuAsnAsnIleAspGlnGlyAla 406
1246 GGAGGATTATATTTCAAGAGATTTTACGTC...TCGCTGAAATAA 1292
407 GlyGlyLeuPhePheGlyLysAspTyrGluValLysGlyThrSerAspAs 423
1293 CGAAACTTGGCAAGCGCGGCTTCATATCAGTGAAGACAGTACGTTA 1342
423 nThrThrTyrLysGlyAlaGlyLysSerValAlaGlyLysThrValT 440
1343 CTGGAAGTAAACGGGTGGCAACGACCGCTGTCCAAATTCGCGAA 1392

440 hTrpLysValHisAsnProGlnTyrAspArgLeuAlaLysIleGlyLys 456
1393 GGCAGCTGCAGCTTCAAGCCAAAGGGAAACCAAGGCTCGATCAGCT 1442
457 GlyThrLeuIleValGlyGlnGlyThrGlyAspAsnLysGlySerLeuLysVal 473
1443 GGGCGAGGTACAGTCAATTTTGCATCGACGCGACAGATTAAGCAAAA 1492
473 lGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGlySerGly...G 489
1493 AACAGCCTTTAGTAATCGCTTGGTGGTGGCGAGGCGGTACGGTGA 1542
489 lHisAlaPheAlaSerValGlyIleValSerGlyArgSerThrLeuVal 505
1543 CTGATGCCCATATATCATGTTCAACCCGCAAACTATTTTCGGCTTTCG 1592
506 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheAr 522
1593 CGGCGAGCTTGGATTTAAAGGCGCATTCGTTTCCACCGATTC 1642
522 gGlyGlyArgLeuAspLeuAsnGlnLysAsnSerLeuThrPheAspHisIleA 539
1643 AAAATACGATGAAGGGCGCATGATGTCACACCAATCAAGCAAAAGA 1692
539 rGAsnIleAspGlnGlyAlaArgLeuValAsnHisSerThrSerLysHis 555
1693 TCCACGCTTACCATTACAGCAATTAAGATATTTGCTACACCGCAATA 1742
556 SerThrValThrIleThrGlyAsp...AsnLeuIleThrAspProAsnAs 571
1743 C..... 1743
571 nValSerIleTyrTyrValLysProLeuGluAspAsnProTyrAlaI 588
1743 1743
588 leArgGlnIleLysTyrGlyTyrGlnLeuTyrPheAsnGlnLysAsnArg 604
1743 1743
605 ThrTyrAlaLeuLysLysAspAlaSerIleArgSerGluPheProG 621
1743 1743
621 nasnArgGlyLysSerAsnAsnSerTrpLeuTyrMetGlyThrGluLysA 638
1744AACAGCTTGATAGCAAAAGCAATT 1770
638 lAspAlaGlnLysAsnAlaMetAsnHisIleAsnAsnGlnArgMetAsn 654
1771 GCCTACAGGCTGTTGGTGGCAAGATACGACCAAAAGCAAGGCGG 1820
655 GlyPheAsnGlyTyrPheGlyGlnGlnGlu...GlyLysAsnAsnGlyLys 670
1821 GCTCAACCTTGTTTACAGCCGCGCAGAGAAGACCGCAACCTGCTGCTT 1870
670 nLeuAsnValIleThrPheLysGlyLysSerGlnGlnAsnArgPheLeuLeu 687
1871 CCGGCGGAACAAATTTAAAGGCAACATCAGCAACAAAGGCAAACTG 1920
687 hGlyGlyThrAsnLeuAsnGlyAspLeuAsnValGlnGlnGlyThrLeu 703
1921 TTTTTCAGGCGAGCAACACCGCATCGCTCAAT..... 1956
704 PheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyLse 720
1957CATTTAAAGCAACATTTGTCGCAAAAGAGGCAATTCCTCGG 1999
720 rSerThrLysLysAspSerHisPheSerGlu.....AsnA 732
2000 GGAATTCGTGTGGCAACAGACTGATCAACCGCATTTAAAGCGGA 2049

732 snGluValValValGluAspAspTPrIleAsnArgAsnPhelYsalThr 748
2050 AACCTCCAAATTAAAGCGGAGACGGCGGTGTTC...CGCAATGTTC 2096
||||| : : : : : ||| |||||
749 AsnIleAsnValIleThrAsnAsnAlaThrIleuTYrSerGlyArgAsnValG1 765
2097 CAATGTAAGGCGGATGGCATTTGGCAATACGCCCAAGCAAGTTTGG 2146
: : : : : ||||| : : : : :
765 uSerIleThrSerAsnIleThrAlaSerAsnAsnAlaLYsValIleHisIleG 782
2147 GGTTCGACCGCATCAAGCCACACATCTGTATCGATTCGGATCGAGAC 2196
|| : : : : : ||| : : : : : ||||| : : : : :
782 IYrYr....LysAlaGlyAspThrValCysValAlaArgSerAspTYrThr 796
2197 GGTCTGACAAATTTGTTCGAAAAAACCATTTACCGAGATTAAGTATGTC 2246
|| : : : : : ||| : : : : : ||||| : : : : :
797 GlyTYrValIleThrCysThrThrAspLYsIleuSer...AspLYsAlaIleuS 812
2247 TTGATTTGACTAAGACCGACATCAAGCGCAATGTGATCTTGGCGATCAG 2296
||||| : : : : : ||||| : : : : :
812 nSerPheAsnProThrAsnIleuArgLYsValAsnIleuThrGluSerA 829
2297 CTGATTTAAATCTCACAGGCGCTTCCACACACTCAAGCGCAATCTTAGTGA 2346
||||| : : : : : ||| ||| : : : : : ||||| : : : : :
829 IAsnPheValIleu...GlyLYsAlaAsnIleuPheGlyThrIleGlnSer 844
2347 AATGGCGATACAGCTTATACAGTACAGCCACACGCCACCCAAAAACGGCAA 2396
: : : : : ||| : : : : : ||||| : : : : :
845 ArgGLYAsnSerGlnValAlaArgIleuThrGluAsn..... 855
2397 CCTTAGCCTGTGGGCATGTCACAGCAACATTATCAAGCCACATTA 2446
855 855
2447 ACGGCACACATCGGCTTCGGGCATGCTTATTTAATAAGCAGCAC 2496
855 855
2497 GCCGTACAAACGGCAGTCTAGCGCTTTCGGCAGCGTAAGCAAACGT 2546
855 855
2547 AAGCATTCGGCACTCAACAGGTATGTCCTCAGCGATTAAGCAGTAT 2596
855 855
2597 TCCATTTGAAGACGCCGCTTACCGGACAAATCAAGCGCGCAGAGAT 2646
855 855
2647 ACGGCATTAACCTTAAAGACAGCGAATGAGCGCTCCGTCAGCGACGGA 2696
||||| : : : : : ||| : : : : :
856SerHisTPrHisIleuThrGlyAsnSerAs 865
2697 ATTAGGAATTTAACTTGACAGCGCCACCATTAACATTCACCTTCCGCT 2746
: : : : : ||||| : : : : : ||||| : : : : :
865 pValHisGlnIleuAspIleuAlaAsnGlyHisIleHisIleuAsnSerAla 882
2747 ATCCGCACGATCGCGCAGGGGGGCAACCGCAGTGGCAGACAGTGGCGG 2796
: : : : :
882 spAsnSerAsn..... 885
2797 CGCGCGCGTGGCGCGCTTCGCGCGCTTCCCTATTCCTGTTACACCGCC 2846
: : : : :
886AsnValThr..... 888
2847 AACCTCGTAGAATCCGTTTCAACAGCTGAGTAACGGCAATTTGA 2896
: : : : : ||||| : : : : : ||||| : : : : :
889LysTYrAsnThrIleuThrValAsn...SerIleuS 899
2897 ACGGTCAAGGAACATTCGCTTTATGTGCGAATCTTCCGCTACCGAGC 2946
: : : : : ||||| : : : : : ||||| : : : : :
899 ergIYAsnGlySerPheTYrTYrIleuThrAspIleuSerAsnLYsGlnGly 915

2947 GACAATTTGAAGCTGGGGAAAGTTCGGAAGGCATTAACCTTGGCGGT 2996
||||| : : : : : ||||| : : : : : ||||| : : : : :
916 AspLYsValValValThrLYsSerAlaThrGlyAsnIleThrIleuGlnVal 932
CAACAATTCGCGCAACACACTGCAACGCTCGAACAATTAAGCGTAGTGG 3046
: : : : : ||||| : : : : : ||||| : : : : :
932 IAlaAspLYsThrGlyIleuPro...AsnHisAsnGlnIleuThrIleuPhe 948
3047 AAGAAAAAGACAAACACCGCTGCGCAAAACCTTAATTTACCCCTGCA 3096
: : : : : ||||| : : : : : ||||| : : : : :
948 spAla.....SerLYsAlaGlnArgAspHisIleuAsnValSerIleuVal 962
3097 AACGACACGTCGATTCGCGCGGTTCGATTCACCACTCAACGCAAGA 3146
: : : : : ||||| : : : : : ||||| : : : : :
963 GlyAsnThrValAspIleuGlyAlaTrpLYsTYrLYsIleuArgAsnValAs 979
CGCGGATTCGCGCTGCAATTCGCGTCAAGAACAA..... 3183
979 nGlyArgTYrAspIleuTYrAsnProGluValGluLYsArgAsnGlnThrV 996
3183 3183
996 AlaSPThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValPro 1012
3184GAGCTTTCGACAAACTCGGCAAG..... 3207
1013 SerValProSerAsnAsnGlnIleuAlaArgValAspGluAlaProVal 1029
3208GCAG 3211
1029 IProProAlaProAlaThrProSerGluThrThrGluThrValAlaG 1046
3212 AAGCCAAAAACAGCGCGGAAAA.....GACACGCGCAAGCCTT 3252
|| : : : : : ||||| : : : : : ||||| : : : : :
1046 IuAsnSerLYsGlnIleuSerLYsThrValGluLYsAsnGluIAspAla 1062
3253 GACGCGCTGATTCGCGCGGCGGATGCGCTGCA..... 3288
1063 ThrGluThrThrAlaGlnAsnArgGluValAlaLYsGluAlaLYsSerAs 1079
3289AAGACAGAAAGCTTGCCGAACCGCGCGGACGAGCGGGGAA 3334
||||| : : : : : ||||| : : : : : ||||| : : : : :
1079 nValLYsAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGluT 1096
3335 ATGTGCGCATTAAGCGCGAG..... 3357
: : : : : ||||| : : : : : ||||| : : : : :
1096 hrLYsGluThrGlnThrGluThrLYsGluThrAlaThrValGluLYs 1112
3358 GAAGAGAAAAACGGGTGCGAGCGGATTAAGACACCGCTTGGCGAAA.. 3405
||||| : : : : : ||||| : : : : : ||||| : : : : :
1113 GlnGluLYsAlaLYsValGluThrGluLYsThrGlnGluValProLYsVal 1129
3406CAGCGCG 3412
1129 IThrSerGlnValSerProLYsGlnGluIleuSerGluThrValGlnProG 1146
3413 AAGCGGAA..... 3420
: : : : : ||||| : : : : : ||||| : : : : :
1146 IAlaGluProAlaArgGluAsnAspProThrValAsnIleLYsGluPro 1162
3420 3420
1163 GlnSerGlnThrAsnThrThrAlaSPThrGluGlnProAlaLYsGluThr 1179
3420 3420
1179 rSerSerAsnValGluGlnProValThrGluSerThrThrValAsnThrG 1196
3421ACCGCGCGGCTTCCACCGCGC 3441
||| ||||| : : : : :
1196 LYsAsnSerValValGluAsnProGluAsnThrThrProAlaThrThrGln 1212


```

3442 TTCGCC.....CGGCGCCG 3455
      |||
1213 ...ProthValAsnSerGluSerSerAsnLysProLysAsnArgHisAr 1228
3456 CGGCGCGCGCGGATTTGGCG...CAACGTGAACCCCAACCGCGAGCCG 3502
      |||
1228 gArSerValArgSerValProHisAsnValGluProAlaThrThrSerS 1245
3503 AACCGCAGCGC.....GACCTGATCAGCGGTTATGCG 3534
      :: |||
1245 eArAsnArgSerThrValAlaLeuGlyAspLeuThrSerThrAsnThr 1261
3535 AATACGCGTTGAGTGAATTTCCGCC.....ACGCTGA 3569
      |||
1262 AsnAlaValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAs 1278
3570 CAGCGTTTCCGCTACAGAGAGATTAGACCGGTA...TTTGCCGAAG 3616
      |||
1278 nValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetLysAsnG 1295
3617 ACCGCGCGAACCGCGTTTGACAAACCGGCATCCGGACACCAACACTAC 3666
      :: |||
1295 IuGlyGlnTyLysnValThrValSerAsnThrSerMetAsnLysAsnTy 1311
3667 CGTTGCGAAGATTTCCGCGCTACCGCGCAACAAACCGCTGCGCAAT 3716
      |||
1312 SerSerSerGlnTyArgArgPheSerSerLysSerThrGlnThrGlnLe 1328
3717 CGGTATGCAAAAACCTCGGAGCGG...CGGTCGCGCATCTGTTT 3763
      |||
1328 uGlyTPRAspGlnThrIleSerAsnAsnValGlnLeuGlyValPheT 1345
3764 CGCAACACCGGCAAAACACCTTGACAGACGATCGGCACTCGGCA 3813
      :: |||
1345 hTyValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThr 1361
3814 CGGCTGCGCAACCGCGCGCTTGCGGCAATCGGCACTGAC...AGGT 3860
      |||
1362 ...LeuAlaGlnValAsnPheTySerLysTyGlyAlaAsnHisTr 1377
3861 CTACATCGGCATCAGCGCGCGGCTTTAGACGCGCGGCTTTGAC 3910
      |||
1377 pTyLeuGlyIleAspLeuGlyTyGlyLysPheGlnSerLysLeuGlnT 1394
3911 ACGGCATCGGAGCAAAATCCGCGCGCTGTCATTCAGCATTCAG 3960
      :: |||
1394 hAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThr 1410
3961 GCASGATCGCGCGCGGCTTGCGGAGATTCGAGATCGAACGCAATCG 4010
      |||
1411 AlaGlyLysAlaPheAsnLeuGlyAsnPheGlyLeuThrProIleValG 1427
4011 CGCAACGCGCATTTGTCGCAAAAACGAGATTACCGCTACGAAACGTCA 4060
      |||
1427 yValArgTySerTyLeuSerAsnAlaAspPheAlaLeuAspGlnAla 1444
4061 AATCGCGACCCCGCGCTTCATTCAAACGCTACCGCGCGCATTAAG 4110
      :: |||
1444 rGlyLeuValAsnProIleSerValLysThrAlaPheAlaGlnValAsp 1460
4111 GCAGATTATTCATCAACCGCGCAACACATTCATCACGCTTATTT 4160
      :: |||
1461 LeuSerTyThrTyHisLeuGlyLys...PheSerValThrProIleLe 1476
4161 GAGCGTGCSTATACGCGCTTCGCGCAAGTCCGGAACGCGCAACGCTCA 4210
      |||
1476 uSerAlaArgTy...AspAlaAsnGlnGlySerGlyLysLysValAla 1492
4211 ATACCGCGCTATTCGTCAGGATTCGCAAAACCGCAGTGCAGATGCG 4260
      |||
1492 snGlyTyArgPheAlaLysValGlnLysGlnGlnGlnTyAsnAla 1508
4261 GCGGTAAACGCGCAATCAAGGTTTCAACGCTGCTCCACGCTGCGCG 4310

```

```

1509 GlyLeuLysLeuLysTyHisAsnValLysLeuSerLeuLeuGlyLe 1525
4311 CGCCAAAGCCCGCACTGCAAGCGCACACACGCGGCGCATTAATGAG 4360
1525 uThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaGlnLeuLys 1542
4361 GCTAC 4365
1542 crphe 1543

```

seq_name: /c9n2.6/plodata/1/1aa/PCTUS_COMB.pep:PCT-US95-10661A-4

seq_documentation_block:

Sequence 4, Application PC/TUS9510661A

GENERAL INFORMATION:

APPLICANT: Washington University, et al.

TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10661A

FILING DATE: 16-AUG-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/296,791

FILING DATE: 25-AUG-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Trecautin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: FP-55941/RFT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1545 amino acids

TYPE: amino acid

TOPOLOGY: unknown

PCT-US95-10661A-4

alignment_scores:

Quality: 1579.50 Length: 1735

Ratio: 1.704 Gaps: 53

Percent Similarity: 53.429 Percent Identity: 26.686

alignment_block:

US-09-303-518D-649 x PCT-US95-10661A-4

Align seg 1/1 to: PCT-US95-10661A-4 from: 1 to: 1545

64 CGCTTCGCTGCTTACTTACGATATGCTGCTGCTGCGATTCCTCC 113

5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaLysAlaLeuThr 21

114 CCAAGCTGGCGGCGCACACTTATTTGCGATCACTACATACATCTATC 163

21 oTyThrGlnAlaAlaLeuValArgAspAspValAspTyGlnIlePheA 38

164 GCGACTTGGCGAAATTAAGCGAAGTTGCGAGCGGCGAGAAATATT 213
|||||
38 rGAsPrheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
|||||
214 GAGGTTTACAAAGGAGGAGTTGTCGCGAAATCAATGCAACAAAG 263
|||||
55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnG 71
264 C...CCGATGATGATTTCTGTGGTGG...CGTAACGGCGTGGCG 307
:|||||
71 yIleProMetIleAspPheSerValAlaAspValAspLysArgIleAla 88
308 CATGGGCGGCGATCAATATATGTCAGCTGCGACAT... 345
:|||||
88 hIleuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
346AACGGCGCTATACAAACCTTGA 368
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnLys 121
369 TTTTGGTGGGAGAGAGAAATCCCGATCAACATCGTTTACTATATAAA 418
121 nAlaLysAlaHis...ArgAspValSerSerGluGluAsnArgTyrTyr 137
419 TTGTGAACCGAATATAT...AACGAGGAGCT 450
:|||||
137 hIleValGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThr 153
451AAGGCCATCCTTATGGGCGGATATGATGATGCGCGGTT 491
154 ThrGluAspGlnAlaGlnLysArgArgGluAspArgTyrTyrMetProArg 170
492 GCATAAATTTGTCACAGATGACAGAACCTGTGA...ATGACCAAGTATA 538
170 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAspSer 187
539 TGGATGGCGGAAATATATGATCAATAATTAACCTGACCGTGTTCGT 588
187 eThrAlaGlyThrTyrAsnAsnLysAspLysTyrProTyrPheValArg 203
589 ATGGGCGAGCGACGCAATAT... 609
204 LeuGlySerGlyThrGlnPheIleTyrGluAsnGlyThrArgTyrGlu 220
610 .TGGCGATCTGATGAGATGAGCCCAATAACCGGAAAGTTCATATCAT 657
220 uTrpLeuGlyLysGlyLysSerAsp...AlaGlyGlyTyrAsnL 236
658ATTGCAAGTCGATCTGTGGCTCGTGGTGGCAATACCTTT 699
236 euLysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrProTyr 252
700 GCACAAATGATGATGAGTGGTGCACAGTCACATTAAGTAGT... 741
253 GluValAsnHisGluAsnAspGlyLeuIleGlyPheGlyAsnSerAsn 269
742GAAAAATTAACATAGCCCATATGCTTTT 772
269 nGluTyrIleAsnProLysGluIleLeuSerLysLysPro...L 283
773 TACACAGCGAGGCTCATTTGGCGACAGTGGCTACCAATGTTATATAT 822
283 euThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyr 299
823 GATGCCCAAGCAAAAGTGTATTAATGAGGCTATTCGAAACGCGCAA 872
300 AspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspTyrTrpAl 316
873 CCCCTATATAGGAAAAAGCAATGCTTCACAGCTGTTGTTAAGATTGCT 922
316 aGlyTyr...AsnLysLysSerTrp 324
923 TCTATGATGAATCTTGTCTGAGATACCCATTCAGTATTCGAAACA 972

324 InGluTrpAsnIleTyrLysProGluPheAlaGlyLysIleTyrGluGln 340
973 CQTCAAAATGGG...AATACCTTTTAACAGCAATTA 1007
341 TyrSerAlaGlySerLeuIleGlySerLysThrAspTyrSerTrpSer 357
1008 TAAATGCGACAGGAAAAATCAATGCGCAACATGACACAAATCTCGCTA 1057
357 rAsnGly... 359
1058 ATGATTAATAACAGCAAGCTT...CAATGTTTAATGTT 1095
360LysThrSerThrIleThrGlyGlyLysSerLeuAsnVal 373
1096 TCTTATCCGAGACAGCAAGAAACCTGTTATCATGCTGAGGTGGT 1145
374 AspLeuAlaAspGlyLysAspLysPro... 382
1146 CAACAGTTATGACCCAGACAGTGAATATGAGAAATATTTCTTATTC 1195
383AsnHisGlyLysSerValThrPhe... 390
1196 ACGAAGAAAGCGAATTTGATTAATACAGCAATATCAAGTCT 1245
391 ..GluGlySerGlyThrLeuThrLeuAsnAsnIleAspGlyGlyAla 406
1246 GAGGATTAATATTTCAAGGAGATTTACGGT...TCGCTGAAATTA 1292
407 GlyGlyLeuPhePheGlyGlyAspTyrGluValLysGlyThrSerAspAs 423
1293 CGAAATTTGCGAGCGCGCGCTCATATACGTGAACAGTACCTGTA 1342
423 nThrThrLysGlyAlaGlyValSerValAlaGluLysThrValT 440
1343 CTGGAAGTAAAGCGCGGCAACGACCGCTGCCAAATGGCGCAA 1392
440 nTrpLysValHisAsnProGlnTyrAspArgLeuAlaLysIleGlyLys 456
1393 GCGAGCTGCGACGTTCAAGCCAAAGGGGAAACCAAGGCTGATCAGCGT 1442
457 GlyThrLeuIleValGluGlyThrGlyLysAsnLysGlySerLeuLysVal 473
1443 GCGCGAGCTTACATGATTTTGGATCAGCAGCGACAGCAATTAAGCAAAA 1492
473 IeGlyAspGlyThrValIleLeuLysGlnIleThrAsnLysSerGly...G 489
1493 AACAGGCTTATGTAATCGGCTGTCAGCGCAGGAGTACGTCGAA 1542
489 nHisIlePheAlaSerValGlyIleValSerGlyArgSerThrLeuVal 505
1543 CTGAATCCGATTAATCACTTAACCCGACAAATCTATTTGCGCTTTCG 1592
506 LeuAsnAspAspLysGluValAlaAspProAsnSerIleTyrPheGlyPheArg 522
1593 CGGCGAGCTTGGATTTAAAGGCAATCGCTGCTGTTCCAGCTATTC 1642
522 gGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPheAsnHisIle 539
1643 AAATACCGATGAAGGCGGATGATGTTCACACCAATCAACAGCAAGAA 1692
539 rGAsnLysArgLysGluAlaArgLeuValAsnHisSerThrSerLysHis 555
1693 TCCACCGTTTACATTAACGCGCAATTAAGATATGCTTACACCGCAATTA 1742
556 SerThrValThrIleThrGlyAsp...AsnLeuIleThrAspProAsnAs 571
1743 C... 1743
571 nValSerIleTyrTyrValLysProLeuGluAspAspAsnProTyrAlaI 588
1743 1743

```
588 leatrglnllelystyglytyrlnleutyrrheasnlnlgnlnasnarg 604
1743 ..... 1743
605 thrttyrlnleulysaspralaserlleatrgserglupheprogl 621
1743 ..... 1743
621 nasnarglgluserasnasertrleutyrmetylthrglulysa 638
1744 ..... AACAGTTGGATAGCAAAAAGAAATT 1770
638 lasprialnlysasnalmetasnlnleasnasnlnlurmetasn 654
1771 GCCTACAAAGCTGGTTGGCGAAGATAGCAACCAAAAGCGAGCG 1820
655 glypheasnnglytyrrhegllyglngluu...glylysasnnglylys 670
1821 GCCTAACCTGTTTACAGCCCGCGAGAAAGCGACCCCTGCTGCTT 1870
670 nleuasnvalthrphelysglylyserglnglnasnargpheuleut 687
1871 CGGGGGAACAAATTAAACGGCAATCAGCAACAAAGCGCAACTG 1920
687 hnglylythrasnleuasnnglylaspleuasnvalglnlgllythrleu 703
1921 TTTTCACGCGAGACCAACACCGCGCTACANT ..... 1956
704 pheleusergllyargprothrprohlsalargspillealagllylese 720
1957 ..... CATTTAAAGACCATTTGGTGGCAAAAAGAGGCAATTCCTCGC 1999
720 tserthrlyslasrserlnlspheuserglu.....AsnA 732
2000 GGAATATGCTGGGCAACGACGTGATCAACGCGACATTTAAAGCGAA 2049
732 snnglvalvalvalglunspasprlpleasnargasnphelyslalthr 748
2050 AACTTCCAATTAAGCGGAGCGAGCGGCTGCTCC...CGCAATGTTGC 2096
749 AsnleasnvalthrAsnAsnAlathrleutysergllyargasnvalgl 765
2097 CAAAGTGAAGCGCATTTGGCATTTGAGCAATCAGCCCAAGCACTTTTG 2146
765 userllethrserasnllthrAlaserasnAsnAlalysvalhlsleg 782
2147 GGTGCGACCGCATCAAGCCACACATCTGTACAGCTTGGAGCTGAGC 2196
782 lytyr.....LysAlaelyasphthrvalCysvalargserasprtyrthr 796
2197 GGTCTGACAAATTTGTGCAAAAAACCATTCAGCGATTAAGTGAATTGC 2246
797 glytyrvalthrCyslythrThrAsprlyseuser...AsprlysalaleuAs 812
2247 TTTCATGACTAAGACCGCATCAGCGCAATGCGATCTCCGATCAGC 2296
812 nserthasnprothrAsnleuArgglYasnValasnleuthrgluserA 829
2297 CTCATTTAAATCTCACAGGCTTGCAACATCAACGGCAATCTAGAGCA 2346
829 lAsnphvalleu...GlylysAlaasnleuphegllythrllleglnser 844
2347 AATGGCGATACAGTTATACAGTCAGCCACCAACGCCCAAAAGCGCAA 2396
845 ArgglYasnserglvalargleuthrglunsn..... 855
2397 CCTTAGCTCTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2446
855 ..... 855
2447 ACGGCACACATCGGCTTCGGCAATGCTTCATTTAATCTAAGCGACAC 2496
855 ..... 855
```

```
2497 GCGGTACAAAACGGCACTCTAGCGCTTCCGGCAACGCTAAGCAAAAGCT 2546
855 ..... 855
2547 AAGCCATTCGCACTCAACGGTAATGTCCTTAGCCGATTAAGCAGAT 2596
855 ..... 855
2597 TCATTTTGAAGACGCCGCTTACCGGCAATACAGCGCGCAAGAT 2646
855 ..... 855
2647 ACGGCATTACACTTAAAGACAGGAATGAGCGCTGCCGTCAGCGACGA 2696
856 ..... SerlnlstrphlsleuthrglylasnserAs 865
2697 ATTAGCAATTTAAACCTTGACAACGCCACCATTTACACTCAATTCGCGCT 2746
865 pvalhlslnleuasnpleuAlasnnglyhlsllenhlsleuasnserAlaA 882
2747 ATGCCCAAGATCGGCGCAGGGCGCAACCGGCAAGTCGACATGCGCGC 2796
882 spasnserAsn..... 885
2797 CGCGCGCGCTTCGGCGCGTTCGCGCGCTTCCCTATTAACGTTACACGCC 2846
886 ..... Asnvalthr..... 888
2847 AACTTCGGTAGAATCCGTTTCAACAGCGGTACGCGTAACGGCAATTGA 2896
889 .....LystyrAsnThrleuthrvalAsn...SerleuAs 899
2897 ACGGTACGGGAACATTCGCTTATATGTCGAACTCTTCGGCTACCGCAGC 2946
899 erglyasnnglyserphetyrtyrleuthrAspleuserasnlysglngly 915
2947 GACAAATTAAGCTGGGCAAGCTTCGAAAGGCATTTACACTTGGCGGT 2996
916 AsprlysalvalvalthrlysserAlathrglyAsnphethrleuAlnA 932
2997 CAACATATCCGCGCAACGACCTGCAAGCCTGCAACAATTCAGCGTAGTG 3046
932 lAlaAsplythrClyglupro...AsnhsAsnngluleuthrleuAspA 948
3047 AAGGAAGAAGCAACAACCGCTCCGAAACCTTATTTACACCTTGCA 3096
948 spAla.....SerlysalaglnargasphlsleuasnvalSerleuval 962
3097 AACGACACGTCGATGCGCGCGGTGCGCTTACCACTATCCGCAAGA 3146
963 GlyasnthrvalAspleuuglyAlatrplystyrylyleuAsnvalAs 979
3147 CGCGAGATTCGCGCTCGATTAATCCGGTCAAGAACA..... 3183
979 nglYasgtyAspleutyAsnproglvalglulysatgAsnnglnthrV 996
3183 ..... 3183
996 alAsprthrAsnllthrThrProasnAsnllleglnAlaAspvalPro 1012
3184 .....GAGCTTCGCAACAATCGCGCAAG..... 3207
1013 SerValproserasnnglnlglleAlaargvalAspAlaAlaprova 1029
3208 .....GCAG 3211
1029 lProproAlaProAlaThrProsergluThrthrngluThrvalAlaag 1046
3212 AAGCAAAAACAGCGGCAAAAA.....GACACCGCGCAAGCGCTT 3252
1046 lunsaserlysglngluserlysthrvalglulysasnnglnlAsnspAla 1062
```

```

3253 GACGCGTATGCGCGCGCGCATCGCTGAA..... 3288
      ||| :|||:|||||:|||||:
1063 ThrGluThrAlaGlnAsnArgGluValAlaLysGluAlaLysSerAs 1079
3289 .... AGACAGAAAAGCGTCCGACACCGCGCGCGCGAGCGGCGGAAA 3334
      |||:|||||: :|||: :|||: :|||: :|||:
1079 nValylAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGlu 1096
3335 ATGCGCGCATTTAGCAGCGGAG..... 3357
      :|||: :|||:
1096 hTrpLysGluThrGlnThrGluThrLysGluThrAlaThrValGluLys 1112
3358 GAAGAGAAAACGCGTGCAGCGCGATTAAGACACCGCTTGCGGAAA... 3405
      |||:|||||: :|||: :|||: :|||: :|||:
1113 GluGluLysAlaLysValGluThrGluLysThrGlnGluValProLysVa 1129
3406 ..... 3412
      :|||: :|||:
1129 LThrSerGlnValSerProLysGlnGluGlnSerGluThrValGlnPro 1146
3413 AAGCGGAA..... 3420
      :|||: :|||:
1146 hAlaGluProAlaArgGluAsnAspProThrValAsnIleLysGluPro 1162
3420 ..... 3420
      :|||: :|||:
1163 GlnSerGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluTh 1179
3420 ..... 3420
      :|||: :|||:
1179 rSerSerAsnValGluGlnProValLThrGlnSerThrThrValAsnThr 1196
3421 ..... ACCGCGCGCGCTACACCGCGC 3441
      :|||: :|||: :|||: :|||:
1196 LysAsnSerValValGluAsnProGluAsnThrThrProAlaThrThrGln 1212
3442 TTCCGC..... 3455
      :|||: :|||:
1213 .. ProThrValAsnSerGluSerSerAsnLysProLysAsnArgHisAr 1228
3456 CCGCGCGCGCGCGATTTGCCG... CAACGCGAACCCGACCGCGAGCCG 3502
      |||:|||||: :|||: :|||: :|||:
1228 gArgSerValArgSerValProHisAsnValGluProAlaThrThrSerS 1245
3503 AACCGCAGCG..... GACGTATCAGCCGTTATGC 3534
      :|||: :|||: :|||: :|||:
1245 eRAsnAspArgSerThrValAlaLeuLysAspLeuThrSerThrAsnThr 1261
3535 AATAGCGGTTGAGTGAATTTCCGC..... ACGTCAA 3369
      :|||: :|||: :|||: :|||:
1262 AsnAlaValLeuSerAspAlaArgAlaGlnPheValAlaLeuAs 1278
3570 CAGCGTTTGGCGGTACAGAGCAATAGACCGGTA... TTGGCGGAG 3616
      :|||: :|||: :|||: :|||:
1278 nValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnSng 1295
3617 ACCGCGCAACGCGCGTTTGACAGACGCGCATCCGGACACCAACACTAC 3666
      :|||: :|||: :|||: :|||:
1295 LuGlyGlnTyAsnValLThrValSerAsnThrSerMetAsnLysAsnTy 1311
3667 CGTTGCGCAAGTTTCCGCGCTACCGCGCAACAAACGCACTCGCGCAAT 3716
      :|||: :|||: :|||: :|||:
1312 SerSerSerGlnTyArgArgPheSerSerLysSerThrGlnThrGlnLe 1328
3717 CGGTATGAGAGAAAACCTCGGACGCGG... CGGTGCGGATCCTGTTT 3763
      :|||: :|||: :|||: :|||:
1328 uGlyLThrPspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPhe 1345
3764 CGCAACAACCGGCAACCAACCTTCGACGACGCGCATCGCAACTCGCA 3813
      :|||: :|||: :|||: :|||:
1345 hTrpValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThr 1361
3814 CGGCTTGCCACGCGCGCTTTTGGGCAATACGCGCATCGAC... AGGTT 3860

```

```

1362 ..... 1377
1362 ...LeuAlaGlnValAsnPheLysSerLysTyTrpAlaAspAsnHisTr 1377
3861 CTACATCGGCATCAGCGCGCGCGGCTTTAGACGCGCGACGCTTCAG 3910
      :|||:|||||: ||| ||| :|||: :|||: :|||:
1377 rPyrLeuGlyIleAspLeuGlyTrpGlyLysPheIleSerLysLeuGlnT 1394
3911 ACGCATCGAGAGCAAAATCCGCGCGCGCTGCTGATTCAGGCAATTCAG 3960
      :|||: :|||: :|||: :|||: :|||:
1394 hAsnAlaAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThr 1410
1411 AlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValG 1427
4011 CGCAACGCGCATTTGCTCCAAAACGCGATTCAGCGGCAACGCA 4060
      :|||: :|||: :|||: :|||: :|||:
1427 yValArgTySerTyLeuSerAsnAlaAspPheAlaLeuAspGlnAla 1444
4061 AATCGCACACCGCGCGCTTCATTCACCGCTACCGCGCGCATTAAG 4110
      :|||: :|||: :|||: :|||: :|||:
1444 rGlyLeuValAsnProIleSerValLysThrAlaPheAlaGlnValAsp 1460
4111 GCAGATTATTCATTAACCGCGCGCAACATTCGATCAGCGCTATT 4160
      :|||: :|||: :|||: :|||: :|||:
1461 LeuSerTyThrTrpHisLeuGlyLys... PheSerValThrProIleLe 1476
4161 GAGCGTGTCTTACCGGATCGCGCTTCGCGCAACGCGCGCA 4210
      :|||: :|||: :|||: :|||: :|||:
1476 uSerAlaArgTy... AspAlaAsnGlnGlySerGlyLysIleAsnVal 1492
4211 AATCGCGGATTTGGCTCAGATTTGGCAAAACCGCGCATGCGGATG 4260
      :|||: :|||: :|||: :|||: :|||:
1492 sngLyTrpAspPheAlaTyAsnValGluAsnGlnGlnThrAla 1508
4261 GCGGTAAACCGCAATCAAGTTTCACGCTGCTCCGCGCGCGC 4310
      :|||: :|||: :|||: :|||: :|||:
1509 GlyLeuLysLeuLysTyHisAsnValLysLeuSerLeuIleGlyGly 1525
4311 CGCAAGCGCGCGCACTGAGACGCAACACGCGCGCGCATCAATTA 4360
      :|||: :|||: :|||: :|||: :|||:
1525 uThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeu 1542
4361 GCTAC 4365
      :|||: :|||:
1542 eRphe 1543
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-296-791-5
seq_documentation_block:
: Sequence 5, Application US/08296791
: Patent NO. 6245337
: GENERAL INFORMATION:
: APPLICANT: St. Gene III, Joseph W.
: TITLE OF INVENTION: Haemophilus Adherence and Penetration
: TITLE OF INVENTION: Protein
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentm Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/296,791
: FILING DATE: 25-AUG-1994

```

```

: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Treacartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (415) 781-1989
:
: TELEFAX: (415) 398-5249
:
: TELE: 910 277299
:
: INFORMATION FOR SEQ ID NO: 5:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1702 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: unknown
:
: US-08-236-791-5

```

```
alignment_scores:
  Quality: 1568.00
  Ratio: 1.673
  Gaps: 49
Percent Similarity: 50.053
Percent Identity: 24.947
```

```
alignment_block:
US-09-303-518D-649 x US-08-296-791-5
.
```

Align seg 1/1 to: US-08-296-791-5 from: 1 to: 1702

[illegible]

```

188  sprlaelgylthrtysanaspaiasnlstysrproalaphavalargyeu 204
592  GGGGACGGACGGCAATATTTGGGAGCTGTGCAAGTGAAG..... 630
205  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
221  glysergylthrlphelethrtysglslaspaanltyrserleu 221
631  .CCCAATTAACCCGGAAGATTCATATCAT.....ATTGCAGCTG 667
221  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
238  eleuasnashnlsiglvalaglyslasnasneulysleuvalaglsapa 238
668  CGTATTCTTGGCTCGTTGGTGGCAATACCTTTGGACAAATGCGATCAGT 717
238  lathrtthrtysgylthlealaglythrprotylglslvalasnlsglsun 254
718  GGGGACGACGTCACCTTAGCTAGTGTGAAAAATAACATACG....CCATA 764
255  Asnlglyleuilellypheelylasneserlysglucglunhlsstasprly 271
765  TGGTTTTTTACCAACAGGA.....GGCTCATTTGGCGACA 799
271  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
280  scglyleleuSerGlnasProleuThrasnTyrlavalLeuGlyasps 288
800  GNGGCTACCAAGTGTATTCATGTATGGCCAAACCAAAAGCGTTATTC 849
288  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
288  erglyserProleuPheValIyraspatrglulysgilylstrleuphe 304
850  AATGGGATTGCMAACGGGCAACCCCATATATAGGAAGCAATGGCTT 899
305  LeuGlySerTyrAspPheThrplaglyltyrAsnLysSertrpGlnGl 321
900  CCAAGCTGGTTCGTAAGAATGGTTCATATGATGAATCTTTGCTGCAATA 949
321  utrrasnlleTyrLysProGluPheAlaLysThrValLeuaspLysaspr 338
950  CCCATTCACATATTCYACAGAACCCGCTCAAAATAGGAATACCTTTTAC 999
338  ||:::||||:::||||:::||||:::||||:::||||:::||||:::
338  hnalellyserleule....glyserAsnThrGlnTyrlAsntrpAsn 352
1000  GACGATTAATATGCGACAGGAAAAATCAATGCCAACATGAACACATTC 1049
353  Pro.....ThrllyLysThrSerVal..... 359
1050  TCTGCCTATTAATTAATAAACAAGAACCGTTCATTTGTTATGTTGCTT 1099
360  ||:::||||:::||||:::||||:::||||:::||||:::||||:::
360  .lleserAsnGlySerGlnSerLeuAsnValAspLeuPheaspserserG 376
1100  TATCCGACAGACGAAGAGAACCTGTTTATCATGCTGACAGTGCTCAAC 1149
376  ||:::||||:::||||:::||||:::||||:::||||:::||||:::
376  lAspThrAspsersTys..... 381
1150  AGTTATCGACCCAGACGATGAATATATGAGAAAAATATTTCTTATTGACGA 1199
382  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
382  .....LysAsnAsnHlsIglYsSerValThrlPhe.....Ar 392
1200  AGGAAAAGGCGAATTTGATCTTACCAGGACGAATCATCATAGAGTCTGAG 1249
392  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
392  gglserSerGlyThrleuThrleuAsnAsnAsnHlspslncgllyadgys 409
1250  GATTTATTTCCAAAGAGATTTTACGGTC...TCGCCGTAATAATACGAA 1296
409  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
409  lYleuPheheglunllyasprtygluValLysGlyThrSerAspsSerthr 425
1297  ACTGTGCAAGGGCGGGCGTTCATTCAGTGAAGACATACCGTTACTGTG 1346
426  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
426  ThrTrpLysGlyAlaIglYalSerValAlaAspGlyLysThrValThrTr 442
1347  GAAAGTAAGGGCTGGCGAAACGACCGCTGTCCAAATTCGGCAAGGCA 1396
442  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
442  rlyValHlsAsnProLysSerAspArgLeuAlaLysIleGlyLysGlyT 459
1397  GCGTCAGCGTTCACCCAAAGGGGAAACCAAGCTGCATACAGCTGGCG 1446
459  |||:::||||:::||||:::||||:::||||:::||||:::||||:::
459  hnleuileAlaIdulYlYsGlyLysAsnLysLysGlySerLeuYalValcyl 475

```

```
1447 GACGTCAGTCAATTTGATCAGACGACGATTAAGCAAAACA 1496
|||||
476 AspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnLysVally 492
|||||
1497 AGCCTTACTGAATCGCTTGTCAGCGGACGGGTACGGTCAACTGA 1546
|||||
492 SalPheSerGlnValGlyIleValSerGlyArgSerThrValIleVal 509
|||||
1547 ATGCGCATATCAGTTCACCCGACAACTCTATTCCGCTTCCGGC 1596
|||||
509 snAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGly 525
|||||
1597 GACGCTTGAATTAAGCGGACGTCCTTCCTCCACCGTATTCAAA 1646
|||||
526 GlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGlnHisIleArgAs 542
|||||
1647 TACCGATGAAGGCGCATGTGTCAACCAATCAAGACAAAGAAATCCA 1696
|||||
542 nIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThrSer 559
|||||
1697 CCGTTACCATTAACAGC 1713
|||||
559 hrValThrIleThrGlyLysSerLeuIleThrAspProAsnThrIleThr 575
|||||
1713 1713
576 ProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheArg 592
|||||
1713 1713
592 GileLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyr 609
|||||
1713 1713
609 yAlaIleuArgLysGlyAlaSerThrArgSerGluLeuProLysAsnSer 625
|||||
1714 1714
626 GlyLysSerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAspAla 642
|||||
1728 TACACCGCGCAATTAACACAGCTTGATACCAAAAAGAAATTCCTACA 1777
|||||
642 aLysArgAsnValIleTasnhIleAsnAsnGluArgMetAsnGlyPhe 659
|||||
1778 ACGGTTGTTGGCGAGAAAGATACGACCAAAAGCAAGCGCGCTCAAC 1827
|||||
659 snGlyTyrPheGlyGluGlu...GlyLysAsnAsnGlyAsnLeuAsn 674
|||||
1828 CTGCTTACCAAGCCCGCGCAAGAACCGCACTGCTGCTTCGCGCGG 1877
|||||
675 ValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuLeuThrGly 691
|||||
1878 AACAATTTAAAGCGCAACATCAAGCAAAACGCAAACTGTTTCA 1927
|||||
691 yThrAsnLeuAsnGlyLysPheLysValGluLysGlyThrLeuPheLys 708
|||||
1928 GCGGACAGCAACACCGCGCTCAAT 1956
|||||
708 ergLysArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThr 724
|||||
1957 CATTTAAACGACCTTGTGCGCAAAAAGAGGCTTCTCGCGGGGAAT 2006
|||||
725 LysLysAspGlnHisPheAlaGlu 736
|||||
2007 CGTGTGGACAAAGCATGTGATCAACGCAATTAAGCGGAAACTTC 2056
|||||
736 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIle 753
|||||
2057 AAATTAAGGCGGACAGCGGTGTTCC...CGCAATGTGTCCAAAGT 2103
|||||
753 snValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIle 769
|||||

2104 AAAGCGATTTGGATTGACAAATCACGCCACAGCTTTTGTGTGCG 2153
|||||
770 ThrSerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr 785
|||||
2154 ACCGATCAAAAGCCACACAAATCTGACACGTTTCGACTGGACGGGTGCA 2203
|||||
786 ...LysAlaGlyAspThrValLysValArgSerAspTyrThrIleTyr 801
|||||
2204 CAATTTGTGTGAAAAAACCATTTACCGGATTAAGTATGTTGTTCAATTG 2253
|||||
801 aThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 816
|||||
2254 ACTAAGACCGACATCAGCGGCAATGTGATCTGCCATCAGCTCATTT 2303
|||||
817 AsnAlaThrAsnValSerGlyAsnValAsnLeuSer 828
|||||
2304 AAATCTCACAGGCTTGCACATCAACGGCAATCTTAGTCAATGGCG 2353
|||||
828 828
2354 ATACAGTTATATACAGTCACGCCAACGCCACCAAAAGCGCACTTAGC 2403
|||||
828 828
2404 CTGCTGGCAATGCCCAAGACAACTTTATATACAGCCACATTAAACGCAA 2453
|||||
829 ...GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThr 843
|||||
2454 CACATCGGCTTGGCGCAATGCTTCATTTAATCTAAGCAGCACGCCGCTAC 2503
|||||
843 rIleSerGlyThrGlyAsnSerGlnValArgLeuThrGlu 856
|||||
2504 AAAACGGCAGTCTGACGCTTCCGCGCAACGCTAAGCAAAACGTAAGCAT 2553
|||||
856 856
2554 TCCGCATCTCAACGGTAATGTCTCCCTACCGCATTAAGCATATTCCATT 2603
|||||
856 856
2604 TGAAGACAGCCGCTTACCGGACAAATCAGCGCGGCAAGATACGCAT 2653
|||||
856 856
2654 TACACTTAAGACAGCGAATGACGCTCCGCTCAGCGACGAAATTAGCC 2703
|||||
857 ...AsnSerHisTrpHisLeuThrGlyAspSerAsnValAsn 869
|||||
2704 AATTTAACTTGAACAACGACCATTTACACTCAATTCGCTATTCGCA 2753
|||||
870 GlnLeuAsnLeuAspLysGlyHisIleHisLeuAsnAla...GlnAs 884
|||||
2754 CGATCGCGGAGGCGCAACCGCGAGTGCACAGATCGCGCGCGCC 2803
|||||
884 nAspAlaAsnLysValThrThr 891
|||||
2804 GTTCGCGCGCTTCCGCGCTTCCCTATTATCCGTTACACCGCAACTCG 2853
|||||
891 891
2854 GTAGATATCCGTTTCAACAGCGTGAAGTAAACGCAATTTGAACGCTCA 2903
|||||
892 ...TyrAsnThrLeuThrValAsn...SerLeuSerGlyAs 903
|||||
2904 GCGAATATCCGCTTATGTGCGAATCTTGCGCTACGCGACGCAAT 2953
|||||
903 nGlySerPheTyrTyrLeuThrAspLysSerAsnLysGlnLysAspLysV 920
|||||
2954 TGAAGCTGCGGAAAGTTCCAGAGCACTTACACTTGGCGGTCAACAT 3003
|||||
920 aLValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAsp 936
|||||
3004 ACCGGCAACGAACTGCAACGCTCGAACAATTGACGAGTAGTGAAGAAA 3053
|||||
```



```

3903 CCTTCAGAGCATGCGAGCAAAATCCGCGCGGTGCTGATTACG 3952
1548 nleuysThAsnHisAlaLysAlaPheAlaArgHisThrAlaGlnPheG 1565
3953 GCATTGAGCAGCATACCGCGCGGTTCGCGGATTCGCGCATGCAACCG 4002
1565 lyeuThrAlaGlyLysAlaPheAsnLeuGlyAsnHegLyleThrPro 1581
4003 CACATGCGCGCAACCGCGCTATTTCGTCACAAAAGCGGATTACCGCTACGA 4052
1582 lLeValGlyValArgTyrSerTyrLeuSerAsnAlaAsnPheAlaLeu 1598
4053 AAAGCTCAATATGCGCCCGCGCGCTTCGATTCACACCGCTACCGCGCG 4102
1598 alysAspArgIleLysValAsnProIleSerValLysThrAlaPheAlaG 1615
4103 GCATTAAAGCAGATTATTCATTCAAACCGCGCAGACACATTCCTCATCAG 4152
1615 lNValAspLeuSerTyrThrTyrHisLeuGlyLys...PheSerValThr 1630
4153 CCTATTATGAGCCGTGCTAT...ACCGATGCGCGCTTCGCGGCAACGCGC 4199
1631 ProIleLeuSerAlaArgTyrAspThrAsnGlnGlySerGlyLysIle 1646
4200 AACACGCGTCATATACGCGCTATTCGCTCAGATTTCGCGCAAAACCGCA 4249
1647 ...AsnValAsnGlnIleTyrAspPheAlaTyrAsnValGlnAsnGlnG 1662
4250 GTGCGGAATGGGCGGTAAACCGCGCAATCAAGGTTTCACGCTGCTCCTC 4299
1662 lNtyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1678
4300 CAGCTGCGCGCGCGCAAGGCGCGCAAGCTGGAAGCGCACACAGCGCGGG 4349
1679 lIleGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaG 1695
4350 CATCAATTAAGGCTAC 4365
1695 nLeuLysLeuSerPhe 1700

seq_name: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:PCT-US95-10661A-5
seq_documentation_block:
; Sequence 5, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-5

alignment_scores:
  Quality: 1568.00      Length: 1872
  Ratio: 1.673          Gaps: 49
  Percent Similarity: 50.053  Percent Identity: 24.947

alignment_block:
US-09-303-518d-649 x PCT-US95-10661A-5
Align seg 1/1 to: PCT-US95-10661A-5 from: 1 to: 1702

64 CGCTTCGCGCGCTTACTTAGCCATATGCGGTGCGCATTCCTTC 113
   ::::::::::: .....:::
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
   ::::::::::: .....:::
114 CCAAGCTGCGCGCGCACACTTATTCGCGCATCACTCAATCAATATC 163
   ::::::::::: .....:::
21 cyrThrGlnAlaAlaLeuValArgAspValAspValAspTyrGlnIlePhe 38
   ::::::::::: .....:::
164 GCGACTTTCGCGCAATTAAGGCAAGTTTGACAGTCGCGCGCGCAAGATAT 213
   ::::::::::: .....:::
38 lGAspPheAlaGlnLysAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
   ::::::::::: .....:::
214 GAGGTTCACAAACAAAGAGGAGTTCGCGCAATCAATCAACAAAGC 263
   ::::::::::: .....:::
55 GluValArgAspLysAsnHisSerLeuGlnAsnValIleuProAsnGln 71
   ::::::::::: .....:::
264 C...CGATGATTTGATTTTCTGTGCTG...CGTACGCGGTGCGG 307
   ::::::::::: .....:::
71 yllePrometIleAspPheSerValAlaAspValAspValAspTyrGlnIleAla 88
   ::::::::::: .....:::
308 CATTCGTCGCGCATCATATATTTGAGCGTGGCCAT...AACGGCGCG 354
   ::::::::::: .....:::
88 hLeuIleAsnProIleTyrValAlaGlyValLysHisValSerAsnGly 104
   ::::::::::: .....:::
355 TATAACAACGTGATTTTGT...GCGGAAGGAAGAA 389
   ::::::::::: .....:::
105 ValSerGlnLeuHisPheLysAsnLeuAsnGlnLysMetAsnAsnGlnLys 121
   ::::::::::: .....:::
390 TCCCGATCAACATCGTTTACT...TATAAATG 421
   ::::::::::: .....:::
121 nAspLysSerHisArgAspValSerSerGlnGlnAsnArgTyrPheSerV 138
   ::::::::::: .....:::
422 TGAACGGAATTAATTAAGCAGGAGCTAAAGC... 456
   ::::::::::: .....:::
138 alGlnLysAsnGlnIleTyrProIleLysLeuAsnGlyLysAlaValThrThr 154
   ::::::::::: .....:::
457 .....CATCCTTATGCGCGCATTAATTCATATGATCCGCGTTTGCA 494
   ::::::::::: .....:::
155 GluAspGlnThrGlnLysArgArgGlnAspTyrTyrMetProArgLeuAs 171
   ::::::::::: .....:::
495 TAAATTTGTCACAGATGCAGAACTGTTGAA...ATGACCAATTATATG 541
   ::::::::::: .....:::
171 pLysPheValThrGlnValAlaProIleGlnLysSerHisSerSera 188
   ::::::::::: .....:::
542 ATGGGCGGAATATATATCGATCAAAATTAATTAACCGCGCTGTCGAT 591
   ::::::::::: .....:::
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
   ::::::::::: .....:::
592 GGGCGACGCGCAATATTCGCGATCTGATGAAGATGAG... 630
   ::::::::::: .....:::
205 GlySerGlyThrGlnPheIleTyrLysGlyAspAsnTyrSerLeuI 221

```

```
631 ..CCAAATAACCGCGAAAGTTCATCAT.....ATGCAAGT 667
|||||.....
221 eteLusnshnHISgluValGlyGlyAsnLeuLysLeuValGlyAspA 238
668 CGTATTCCTGGCTCGTGGTGGCAATACCTTGGCAAAATGATGACAGT 717
|||||.....
238 latYrThrTyrGlyLeuAlaGlyThrProTyrLysValAsnHISgluSn 254
718 GGTGGCACAGTCACTTAGTAGTAAAAAATTAACATAGC...CCATA 764
|||||.....
255 AsnGlyLeuIleGlyPheGlyAsnSerLysgluHISserAspProLy 271
765 TGGTTTTTACCAACAGCA.....GGCTCATTTGGCGCA 799
|||||.....
271 sglYleuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288
800 GTGGCTCACCAATGTTTATCTATGATGCCAAAGCAAAAGTGTTAAT 849
|||||.....
288 erGlySerProLeuPheValTyrAspArgGlyLysGlyLysTrpLeuPhe 304
850 AATGGGTATTCGAACGGGCAACCCCTATATAGAAAAAACCAATGCTT 899
|||||.....
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnG 321
900 CCAGCTGGCTGTAAGATGGTTCTATGATGAATCTTGTCTGGAGATA 949
|||||.....
321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAsp 338
950 CCCATTGCGATTCTACGACACCGCTCAAAATGGAAATACTCTTTTAA 999
|||||.....
338 hrAlaGlySerLeuIle.....GlySerAsnThrGlnTyrAsnTrpSn 352
1000 GACGATATTAATGGCACGAAATAATGCCAAACATGAACCAATTC 1049
|||||.....
353 Pro.....ThrGlyLysThrSerVal..... 359
1050 TCTGCTTAATAGATTAAACACGACGCTCAATGTTTAAATGTTTCT 1099
|||||.....
360 ..IleSerAsnGlySerGlnSerLeuAsnValAspLeuPheAspSerSer 376
1100 TATCGAACAACGAAAGAACCTGTTATCTGCTGCGAGTGTGTCAAC 1149
|||||.....
376 InAspThrAspSerLys..... 381
1150 AGTTATCGACCCAGACTGATATATGAGAAATATTTCTTATTATGACA 1199
|||||.....
382 .....LysAsnAsnHISglYlysSerValThrLeu.....Ar 392
1200 AGGAAAGGCGAATTGATCTTACCAGACATCAATCAAGCTGCTGAG 1249
|||||.....
392 gclYserGlyThrLeuThrLeuAsnAsnIleAspGlnGlyAlaGly 409
1250 GATTATATTTCCAGAGGATTTTACGGTC...TGGCTGAAATTAACGAA 1296
|||||.....
409 LysLeuPhePheGluGlyAspTyrGluValLysGlyThrAspSerSerThr 425
1297 ACTTGGCAAGGCGCGGCTTCATATCAGTAAGACAGTACCTGTTACTG 1346
|||||.....
426 ThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysThrValThrTr 442
1347 GAAAGTAAACGGCGTGCAGAACGACCCCTGTCAGAAATCGGCAAGCA 1396
|||||.....
442 PLYsValHISAsnProLysSerAspArgLeuAlaLysIleGlyLysGly 459
1397 CGGTGACGCTCAAGCGCAAGGGGAAAAACCAAGCTCGATCAGCGTGGC 1446
|||||.....
459 hrLeuIleValGluGlyLysGlyLysLysLysGlySerLeuLysValGly 475
1447 GACGATACAGTCAATTTGATCAGACGACGACGATTAAGGCAAAACAA 1496
|||||.....
476 AspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysVal 492
1497 AGCCTTAGTGAATTCGCTTGGTGCACCGCGACGGGTACGTTCAACTGA 1546
|||||.....
492 salAPheSerGlnValGlyLeuValSerGlyArgSerThrValValLeuA 509
1547 ATGCCGTAATTCAGTTCAACCCGACAACTATTTGGCTTGGCGGCG 1596
|||||.....
509 snAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGly 525
1597 GACGTTTGGATTTAAACGGGCAATTCGCTTCCGTCACCGTATTCAAA 1646
|||||.....
526 GlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGluHISleArgAs 542
1647 TACCGATGAAGGGCGGATGATGTGCAACCAATCAAGACAAAGATCCA 1696
|||||.....
542 nIleAspAspGlyAlaArgLeuValAsnHISAsnThrSerLysThrSer 559
1697 CCGTTACCTTACAGCG..... 1713
559 hrValThrIleThrGlnGlyGlySerLeuIleThrAspProAsnThrIleThr 575
1713 ..... 1713
576 ProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheArgAr 592
1713 ..... 1713
592 gIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyrT 609
1713 ..... 1713
609 yAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAsnSer 625
1714 .....AATAAGATATTGC 1727
626 GlyGlySerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAspAlaAl 642
1728 TACAACCGCAATTAACACACGCTTGATGCAAAAGAAATTCCTTACA 1777
|||||.....
642 alYsAlaGlyAsnValMetAsnHISleAsnAsnGlnAlaGlyMetAsnGlyPheA 659
1778 AGCGTTGGTTGGCGAGAAAGATACGACCAAAACGAGCGGCGCTCAAC 1827
|||||.....
659 snGlyTyrPheGlyGlnGluGlu...GlyLysAsnAsnGlyLysLeuAsn 674
1828 CTGTTTACCGCCCGCCGAGAGAACCGCACCTCTCTCTTCCGCGCG 1877
|||||.....
675 ValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuLeuThrGlyG 691
1878 AACAAATTTAAACGCAATCACGCAACAAACGCGCAAACTGTTTTC 1927
|||||.....
691 yThrAsnLeuAsnGlyAspLeuLysValGlyLysGlyThrLeuPheLeuS 708
1928 GCGGACGACCAACCGCACGCGCTTACAAT..... 1956
|||||.....
708 erGlyArgProThrProHISAlaArgAspIleAlaGlyIleSerSerThr 724
1957 CATTTAAGACCATGGTGGCAAAAGAGGGGATTCCTCGGGGGCAAT 2006
|||||.....
725 LysLysAspGlnHISPheAlaGlu.....AsnAsnGluVal 736
2007 CGTGTGGGCAACGACTGATGATCAACCGCATTAATTAAGCGGAAACTTC 2056
|||||.....
736 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleA 753
2057 AAATTAAGCGGACAGCGCGGTGTTTC...CGCAATGTTCCAAAGTG 2103
|||||.....
753 snValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIle 769
2104 AAGCGGATTTGGCATTTGAGCAATCACGCCCAAGCAAGCTTTGGGTGCG 2153
|||||.....
770 ThrSerAsnIleThrAlaSerAspAsnAlaLysValHISIleGlyTyr.. 785
2154 ACCGATCAAAAGCCACCAATCTGTACAGTTGCGAGCTGGAGGCTCTGA 2203
```

[illegible]

952	...	serAsnAlaThrArgAsnAspLeuAsnValSerLeuValGlyAsnT	967
3104	ACGTGATGCGCGCGCGGCGGTACCAACTCATCCGCAAGACGGCGAG	3153	
967	hValAspLeuIleValAlaPlystyryLysLeuArgAsnValAsnGlyArg	983	
3154	TTCCGCGCTGCATTAATCCG...	3171	
984	TYrAspLeuTYrAsnProGluValGluLysArgAsnGlnThrValAspTh	1000	
3171	...	3171	
1000	rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP	1017	
3171	...	3171	
1017	roSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro	1033	
3171	...	3171	
1034	AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy	1050	
3171	...	3171	
1050	sGlnGluSerLyThrValGluLysAsnGluGlnAspAlaThrGluThrT	1067	
3171	...	3171	
1067	hValAlaGlnAsnGluValAlaGluGluAlaLysProSerValLysAla	1083	
3171	...	3171	
1084	AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh	1100	
3172	...	3208	
1100	rglnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLysA	1117	
3209	CAGAACGCCAAAAACAG...	3225	
1117	lAlaValGluLysGluGluLysAlaLysValGluLysAspGluIleGln	1133	
3225	...	3225	
1134	GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl	1150	
3225	...	3225	
1150	aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA	1167	
3225	...	3225	
1167	lGlnProGlnThrGlnSerThrValAlaAlaAlaGluAlaThrSer	1183	
3226	...	3242	
1184	ProAsnSerLyProAlaGluGluThrGlnProSerGluLysThrAsnAl	1200	
3243	GCAAAACCTTGACGCCGCTGATTGGC...	3267	
1200	agluProValThrProValValSerLysAsnGlnThrGluAsnThrTra	1217	
3268	...	3294	
1217	spGlnProThrGluArgGluLysThrAlaLysValGluThrGluLysThr	1233	
3295	GAAAGCGTTGCCGACCGCGCGCAGCAGAGC...	3327	
1234	GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluGlnSe	1250	
3328	...	3343	
1250	rgluThrValGlnProGlnAlaValLeuGluSerGluAsnValProThrV	1267	

```

3344 TATATGAGCGGAGAGAGAAAAAGGCTGACGGCGATTAAGAC... 3390
      ::::::::::::::::::::
1267 AAlasmsnlaGluglValGlnlaGlnleuGlnthGlnthSerAla 1283
3390 .....
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300
3391 ...ACGGCTTGCGGAAACACAGCGGAGCGGAAACCGCGCGGTACC. 3435
      |||||
1300 rAlaThrAlaIleThrGlnThrAlaGlnLysSerAspLysProGlnThrG 1317
3435 .....
1317 LuThrAlaAlaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333
3436 .....ACGGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3455
      |||||
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerSerGluProLysSerAr 1350
3456 CCGCGCGCGCGCGGATTTGCCGCACTGCA... 3486
      |||||
1350 gArg...ArgArgSerIleSerGlnProGlnLysThrSerAlaGlnGluT 1366
3486 .....
1366 hrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382
3487 CCGCAACGCGACCGCCCAACCGCGAC... 3510
      |||||
1387 SerLysProAsnArgArgSerArgArgSerValArgSerGluProThrVa 1399
3511 .....CGCGACCTGATCAGCC 3526
      |||||
1399 LThrAsnGlySerAspArgSerThrValAlaIleuArgAspLeuThrSerT 1416
3527 GTTATGCCAATAGCGCTTGAGTGAATTTCCGCC... 3561
      ::::::::::::::::::::
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432
3562 ACGTCACAGCGGTTTCCGCTACAGACGCAATTAACCGCGCTA...TT 3608
      ::::::::::::::::::::
1433 AlaLeuAsnValGlyLysAlaValaSerGlnHisSerGlnleuGluMe 1449
3609 TGCCGACAGCGCGCGCGCGCTTGACAAAGCGCATCCGCGGACACA 3658
      ::::::::::::::::::::
1449 LAsnAsnGlnGluGlnThrAsnValThrValSerAsnThrSerMetAsnG 1466
3659 AACACTACCGTTCCGACAGATTTCCGCGCTACCGCGCAACAACCGACCTG 3708
      ::::::::::::::::::::
1466 LuAsnLysSerSerSerGlnThrArgArgPheSerLysSerThrGln 1482
3709 CGCCAAATCGGATGACAGAAAACCTCGGACGCGG...CGCGTCGCGAT 3755
      |||||
1483 ThrGlnleuGlyTTPAspGlnThrIleSerAsnAsnValGlnleuGlyG1 1499
3756 CCGTTTTCGACAAACCGACCGGAAACACCTTGACAGCGCATCCGCA 3805
      ::::::::::::::::::::
1499 yValPheThrValaArgAsnSerAsnAsnPheAspLysAlaSerSerL 1516
3806 ACTCGGACGCGCTTGCGCGCGCGCGCTTGCGGCAATACGCGATCGAC 3855
      ::::::::::::::::::::
1516 yAsnThr...LeuAlaGlnValaAsnPheLysThrLysTyrTyrAlaAsp 1531
3856 ...AGGTTCTACATCGGATCAGCGCGCGCGCGCTTTTACAGCGGACG 3902
      ::::::::::::::::::::
1532 AsnHisThrLysGlyLysAspLeuGlyTyrGlyLysPheGlnSerAs 1548
3903 CCTTTCACAGCGCATCGGAGGCAAAATCCGCGCGCTGCTGCATTACG 3952
      ::::::::::::::::::::
1548 nleuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheG 1565

```

```

3953 GCATTTCAGGACGATATACCGCGCGCTTGCGCGGATTCGGCATCGAACC 4002
      ||:::
1565 LyeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrPro 1581
4003 CACATCGCGCGCAACGCGCTATTTCGTCAAAACCGGATTCACGCTACGA 4052
      ::::::::::::::::::::
1582 lLeValGlyValaArgTyrSerTyrLeuSerAsnHisPheAlaLeuAl 1598
4053 AAACGTCATATCGGCAACCGCGCGCTTGCAATTCACCGCTACCGCGCG 4102
      ::::::::::::::::::::
1598 aLysAspArgIleLysValaAsnProIleSerValLysThrAlaPheAla 1615
4103 GCATTAAAGCGCATATTCATTCATAACCGCGCGCAACGATTCGATCAGC 4152
      ::::::::::::::::::::
1615 lNValaPheLysSerTyrThrThrHisLeuGlyGlu...PheSerValThr 1630
4153 CCTTATTGAGCTGCTGCTAT...ACGATGCGCGCTTGCGGCAAGTCCG 4199
      |||||
1631 ProIleLeuSerAlaArgTyrAspThrAsnGlnGlySerGlyLysIle.. 1646
4200 AACACGCGGTCATACCGCGCGTATTCGTCAGATTCGCGCAAAACCGCA 4249
      ::::::::::::::::::::
1647 ...AsnValaAsnGlnTyrAspPheAlaTyrAsnValaGluAsnGlnGlnG 1662
4250 GTGCGGAATGGGCGGTAAACGCGGAAATCAAGATTTCACGCTGCTCCG 4299
      ::::::::::::::::::::
1662 lNtyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1678
4300 CACGCTGCGCGCGCGCAAGCGCGCAACTGGAAGCGCAACACGCGCGG 4349
      ::::::::::::::::::::
1679 lLeuGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaG1 1695
4350 CATCAATTAAGCTTAC 4365
1695 uleuLysLeuSerPhe 1700

```

seq_name: /cgn2-6/plodata/1/1aa/backfiles1.pep:5268270-2

seq_documentation_block:

```

; Patent No. 5268270
; APPLICANT: Meyer, Thomas F.; Halter, Roman; Pohlner, Johannes
; TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM
; NEGATIVE HOST CELLS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171,872
; FILING DATE: 01-JUL-1987
; SEQ ID NO:2:
; LENGTH: 1507
; 5268270-2

```

alignment_scores:

```

Quality: 1532.50 Length: 1709
Ratio: 1.770 Gaps: 43
Percent Similarity: 50.673 Percent Identity: 27.092

```

alignment_block:

US-09-303-518D-649 x 5268270-2 ..

Align seg 1/1 to: 5268270-2 from: 1 to: 1507

```

49 AAACGCGCGCGCATCGCGCTTGCTGCTTACTTGGCATATGCTGTC 98
      ||:::
2 LysAlaLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGCGCATTTCCCAAGCGCTGCGGAGACACACTTATTTCGCGCAT 148
      ::::::::::::::::::::
18 aTyralaLeuThrProTyrSerGlnAlaAlaLeuValaLArgAspArgVala 35
149 ACtACCAATATCTATCGGACTTTCGCGGAAATTAAGGCAAGTTTGACATC 198
      ::::::::::::::::::::
35 sPrtyGlnIlePheArgAspPheAlaGlnLysGlyLysPhePheVal 51

```



```
591 IlePheMetClyTYrThrGlnGlnGluAlaArgLysAsnAlaMetAsnH1 607
1777 .....AACGTTGGCTTTGGCGGGAAG 1798
607 sLysAsnAsnAlaArgTlLeGlyAspGluGlyPhePheAspGlnGluA 624
1799 ATAGCAACAAAGCAAGCGGCGCTCAACCTGTTTACCAGCCGCCGCA 1848
624 sngLysGlnGlnHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSer 640
1849 GAAGACCCGACCCCTGCTGCTTCCGCGGAACAAATTTAAACGGCAACAT 1898
641 AlaGlnAsnArgPheLeuLeuThrGlyGlyAlaAsnLeuAsnGlyLysI1 657
1899 CACGCAACAAACGCAACCTGTTTTTACGCGGAGCAACCAACCCGACG 1948
657 eSerValThrGlnGlnGlyAsnValLeuLeuSerGlyArgProThrProH1SA 674
1949 CCTACATCATTTTAAACGACACCATTTGTCGCAAAAAGAGGC.....ATT 1992
674 laArgAspPheValAsnLysSerSerAlaArgLysAspAlaH1HisPheSer 690
1993 CCTCGCGGGGAATGCTGTGGACACACATGATCAACGCGACATTTAA 2042
691 LysAsnAsnGluValValPheGlnAspArgPheAsnArgThrPheLys 707
2043 ACCGCAAACTTCCAAATTTAAAGCGGACAGCGGTGTTCC...CGCA 2089
707 sAlaLeuGlnIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgA 724
2090 AAGTTGCCAAAGTGAAGCGCATTTGGCATTTGACCAATCACGCCCAAGCA 2139
724 sNValAspAspIleThrAlaAsnIleThrAlaThrAspAsnAlaLysVal 740
2140 GTTTTGGGTGCGACCGCATCAAAAGCCACACATCTGTACAGCTTGGGA 2189
741 AsnLeuGlyTYr.....LysAsnGlyAspGluValLysValArgSerAs 755
2190 CTGGACGGGTCTGCAAAATGTGTGCAAAAACATTAACGCGATTAAG 2239
755 pLysThrGlyTYrValThrCysAsnThrGlyAsnLeuSerAspLysAlaL 772
2240 TGATTGCTTCATGTAGTAAAGCCGACATCAGCGCAATGTGATCTTGCC 2289
772 euAsnSerPheAsp**AlaThrLysIleAsnGlyAsnValAsnLeu... 787
2290 GATCAGCGCTCATTTAAATCTCACAGGCTTCCACACTCAACGCGCATCT 2339
787 ..... 787
2340 TAGTCAAAATGGCGATACAGCTTATACAGTCACACAGCCACCCCAA 2389
788 .....A 788
2390 ACGCAACCTTAGGCTGCTGGGCAATGCCCAAGCAATTTAATCAAGCC 2439
788 sngInAsnAlaIleAlaLeuVal.....LeuGlyLysAla 798
2440 ACATTTAAACGGCAACATCGGCTCGGCAATGCTTCATTTAATCAAG 2489
799 AlaLeuThrPglyLysIleGlnGlyGlnGlyLysSerArgValSerLeuAs 815
2490 CGACACGCCCGTACAAAAGCGCACTGACGCTTTCGCGCAACGCTAAG 2539
815 nGlnHis..... 817
2540 CAAACGTAAAGCATTCGCACTCAACGTAATGTCCTAGCCGATTAAG 2589
817 ..... 817
2590 GCAGATTTCATTTTGAAGACGCGCTTTACCGCAAAATCAAGCGCGG 2639
817 ..... 817

817 ..... 817
2640 CAAGATACGCGATTAACCTTAAAGACAGCAATGACGCTGCCCTAG 2689
818 .....SerLysThrPheIleuThrGlyA 825
2690 GCACGGAATTTAGCAATTTAAACCTTGACAAAGCCACCATTTACSTCAAT 2739
825 sPserGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsn 841
2740 TCCGCTATCGCCAGATGCGGACAGGCGCAACACCGCACTGGGACAGA 2789
842 AsnAla.....SerAspAlaGlnSerAla..... 849
2790 TCGCGCGCGCGCGCTTCCGCGCGTTTCGCGCGTTCCCTATTATCCGTTA 2839
849 ..... 849
2840 CACCGCAACTTCGGTAGATCCCGTTTCAACACGCTGACGCTAAGCGC 2889
850 .....AsnLysTYrHisThrIleLysIleAsn... 858
2890 AAATTCACGCTCAGGAAACATTCGCTTTATGTGCAACCTTCGCGCTA 2939
859 HisLeuSerGlyAsnGlyHisPheHisTYrLeuThrAspLeuAlaLysAs 875
2940 CCGCAGCCACAATTTGAAGCTGCGGGAAGTTCCGAAGGCACTTACACT 2989
875 nLeuGlyAspLysValLeuValLysGluAlaSerAlaGlyHisTYrGlnL 892
2990 TGGCGTCAACATATCCGCAACGCAACCTGCAACGCTCGAACAATTTACG 3039
892 euHisValGlnAsnLysThrGlyGlnPro...AsnGlnGlnGlyLeuAsp 907
3040 GTATGGAAGGAAGAACACAAACCCGCTGCGAACAACCTTAATTTAC 3089
908 LeuPheAspAlaSerSerValGlnAspArgSer...ArgLeuPheValSe 923
3090 CCTGCAAAACGACACGTCGATGCGCGCGCTGCGCTTACCAACTCATTC 3139
923 rLeuAlaAsnHisTYrValAspLeuGlyAlaLeuArgTYrThrIleLysT 940
3140 GCAAAGACGGGAGTTCGCTGCTGATAAT..... 3168
940 nGlnLysnGlyIleThrArgLeuTYrAsnProTYrAlaGlyAsnGlyArg 956
3169 CCGGTCAAGAACAAGACTTCCGACAAACTGCGCAGGCAAGCAACCAA 3218
957 ProValLysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaG 973
3219 AAACACAGCGGAAAGAACACAGCAAGCGCTTGACGCGCTGATTGGCG 3268
973 nLysAlaThrGlnThrAspGlyAlaGln.....IleAlaL 985
3269 CCGGCGCGGATGCGCTC.....GAAAGACGGAAGACGTT 3303
985 ysProGlnAsnIleValValAlaProProSerProGlnAlaAsnGlnAla 1001
3304 GCCGACCGCGCGCGGACGCAAGCGGGGAAATGTGCGCTTTGACAGC 3353
1002 GlnGluLeuArgGlnGlnAlaLysAlaGlnGlnValLysArgGlnInM1 1018
3354 G.....GAGGAAGAGAAAGAACGGG 3373
1018 aAlaGluAlaGlnLysValAlaArgGlnLysAspGlnGlnAlaLysArgL 1035
3374 TGCAGCGCGAT..... 3384
1035 ysAlaAlaGlnIleAlaArgGlnGlnGlnGlnAlaArgLysAlaAlaGlu 1051
3385 .....AAGACACCGCGCTTGGC 3401
1052 LeuAlaLysGlnLysAlaGlnAlaGlnArgLysAlaArgGlnLeuAl 1068
```



```

;      LENGTH: 1848 amino acids
;      TYPE: amino acid
;      TOPOLOGY: unknown
US-08-296-791-6

```

Quality:	1486.00	Length:	1916
Ratio:	1.490	Gaps:	58
Percent Similarity:	52.035	Percent Identity:	24.843

US-09-303-518D-649 X US-08-296-791-6

Align seg 1/1 to: US-08-296-791-6 from: 1 to: 1848

5
LysPheLysLeuAsnPheIleAlaLeuThrValAlaIleArgLysAlaLeuThrPr 21

114 CCAAGCCGGGGGGGACACACTTATTTGGCATCTACCAATACATATC 163

21 otyrThrIleAlaLeuValAlaGAspAspValAspLysGlnIlePheA 38

164 GCGACTTTGCCGAATAATAAGCAAGTTTGAGTCGGGGGGAAGAATAT 21 3

38 rGaSPheAlaIGlnAsnLysSelLysPheSerValGlyAlaThrAsnVal 54

214 GAGGTTTACAACAAAAAGGGAGCTGTGCGCAATCAATGACAAAAGC 263

55 GluValAlaGAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71

264 C..CCGATGATTTATTTTCTGTGGTGTGC..CGTAAGCGCGTGGCGG 307

71 yIleProMetIleSPheSerValValAspValAspLysArgIleAlaI 88

308 CATTTGGTGGCGATCAATATATTTGTGACGCGTGACAT...AAGCGGCG 354

88 hrLeuValAsnProGlnIleValValGlyAlaLysThrSAlaSerAsnGly 104

355 TATAACAACGTTGATTTTGGT.....GCGGAGAGGAAGA 369

105 ValSerGlnLeuHisPheGlnAsnLeuAsnGlnLysAsnMetAsnAsnGlyAs 121

390 TCCGCGATCAACATCGTTTACT.....TATAAATTG 421

121 naLysSerHisArgAspValSerSerGlnGlnAsnArgIleThrThv 138

422 TGAACGGAATATTAAT.....AAGCA 444

138 alGlnLysAsnAsnPheProThrGlnAsnValThrSerPheThrLysGln 154

445 GGGACTTAAGGCATCCTTAAGCGGCGATTAATCATATGCCGCGTTTGA 494

155 GluGlnAspAlaGlnLysArgArgGlnAspLysIleThrProArgLeuAs 171

495 TAAATTTGTCAACATCCAGAACCTGTTGAATGACAGATTAATAGATG 544

171 pLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaAsnAsnA 188

545 GGGCG...AAATATATGCATCAATAATTAACCTGACGCTGTTGCTATT 591

188 snLysGlnGlnLysArgAsnSerAspLysIleProAlaPheValArgLeu 204

592 GGGCGAGGACGGCAATAT.....TG 611

205 GlySerLeuThrGlnPheIleTyrLysGlySerArgTyrGlnLeuI 221

612 GCGATTCGATGAAGATAGCCCATATACCCGGAAGTTTCATATCAT... 657

221 eleuThrGlnLysAspLysGlnGlnAsnLeuLeuArgAsnThrAspAlaG 238

658ATTGCAAGTCCGTAATTCCTTGCGCTGCTGGT 667

```

238  LgelyspsasnlenuglueuValGlyAsnAlaIlyThrTylGlyIleLea 254
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
688  GGCAATACCTTTTGCACAAATAGATGATGGGCGACAGTCACTAAAG 737
      |||::: :::: ||| ::::::::::::::|::| ::::::::::::::|
255  GlyThrProTylIysValAsnHisgluAsnAsnGlyLeuIleGlyPheG 271
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
738  TACTGAAAAAATTAACATAGC...CCATATGGTTTTTACCAACAGGA 783
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
271  yAnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 288
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
784  .....GGCTATTTGGCGACAGGGCGACCAATGCTTTATC 819
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
288  roLeuThrAsnTylAlaValLeuGlyAspSerGlySerProLeuPheVal 304
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
820  TATGATGCCCAAAAGCCAAAGCTGTAAATTAATGGGGATTTGCAACGG 869
      ||||| ::::|::|::|::|::|::|::|::|::|::|::|::|
305  TylAspArgGluGlySerLysThrPheLeuGlySerTylAspPheTr 321
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
870  CAACCCCTTTATAGCAAAAGCAATGGCTCCACGGTGGTGAAGAATT 919
      ||| ::::|::|::|::|::|::|::|::|::|::|::|::|
321  palaglyTylAsnLysLysSerTrpGlnGluTrpAsnIleTylLysHisG 338
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
920  GGTTCATGATCAAAATCTTTCCTGGAGATACCATTACATATTCACGAA 969
      ||| ::::|::|::|::|::|::|::|::|::|::|::|::|
338  LurPheIleGluLysIleTyr.....GlnGlnTyrSerAlaGlySerLeu 352
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
970  CCACGCTCAAAATGGCAAAATACCTTTTAAACAGCAGATAATATGCGACAG 1019
      ::::|::|::|::|::|::|::|::|::|::|::|::|
353  IleGlySerAsnThrGlnTyrThrTrp.....GlnAlaThrGln 365
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1020  AAAAATCATATGCCAAACATGACACAAATCTCTGCTAATAGATTAAAA 1069
      | ::::|::|::|::|::|::|::|::|::|::|::|::|
365  y.....Set 367
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1070  CACGAACCGTT.....CAATTGTTTAATGTTTCTTTATCCGAG 1107
      ||| ::::|::|::|::|::|::|::|::|::|::|::|::|
367  hrSerThrIleThrGlyGlyGluProLeuSerValAspPheThrAsp 383
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1108  ACACGACAGGAACCGTTTATATCATCTGACAGGTGTCACACGTATACG 1157
      ::::|::|::|::|::|::|::|::|::|::|::|::|
384  GlyLysAspLysPro..... 388
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1158  ACCCAAGCTGAATTAATGAGAAATATTTCTTTATACGACGAGAAAG 1207
      |||:::|::|::|::|::|::|::|::|::|::|::|::|
389  .....AsnHisGlySerLethrLeu.....LysGlySerG 400
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1208  GCGAATGATTACTTACACGACATCATCAAGGTGCTGGAGATTATAT 1257
      |||::|::|::|::|::|::|::|::|::|::|::|::|
400  LylThrLeuThrLeuAsnHisIleAspGlnGlyAlaGlyLeuPhe 416
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1258  TTTCGAAGGAGATTTAACGTC...TCGGCTGAAATTAACGAACCTTGCA 1304
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
417  PheGluGlyAspTylGluValLysGlyThrSerAspSerThrThrTyr 433
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1305  AGCGCGGGCGCTTACATATCAGTAAGACAGTACCGTCTACTGGAAAGTAA 1354
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
433  scGlyIleGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1355  AGCGGCTGGCAACGACCGCCCTGTCCAAATTCGCAAGGACGCTGCAC 1404
      ::::|::|::|::|::|::|::|::|::|::|::|::|
450  lAsnTrpLysTylAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1405  GTTTCAAAGCCAAAGGGGAAACCAAGAGCTGCATACGACGGGCGACGGTAC 1454
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
467  ValGluGlyLysGlyLysAsnGluLylLeuLeuLysValGlyAspGlyTh 483
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1455  AGCTATTTTGGATCAGCAGCGCAGCAGATTAAGCAAAAAACAAGCCTTTA 1504
      |||||::|::|::|::|::|::|::|::|::|::|::|::|
483  rValIleLeuLysGlnLysAlaAspAlaAsnLysValGlnAlaPheS 500
      ::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1505  GTGAAATCGCCTTGGTCAGCGCGCAGGGGTACGGTGCACATCCGAT 1554
      ||:::|::|::|::|::|::|::|::|::|::|::|::|

```

500 ercInValIGlyIleValSerGlyArgSerThrLeuValLeuAsnAsp 516
 1555 AATCAGTCAACCCGCAAACTATTCGGTTCGGCGGAGCTT 1604
 517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyValArg 533
 1605 GGATTTAAAGCGCATTCGCTTCGTTCCACCGTATTCAAATACGAG 1654
 533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAsp 550
 1655 AAGGCGCATGATGTTCAACCAATCAAGACAAAGAAATCCACCTTAC 1704
 550 spGlyIaArgValIaAsnHisAsnMetThrAsnThrSerAsnIleThr 566
 1705 ATTCACAGC..... 1713
 567 IleThrGlySerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
 1713 1713
 583 nIleGluAlaGlnAspAspHisProLeuArgIleArgSerIleProt 600
 1713 1713
 600 YrArgIleuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeuLys 616
 1713 1713
 617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerHis 633
 1714 1738
 633 nGluAsnTrpLeuTyrMetGlyArgThrSerAspAlaIaIaLysArgAsnV 650
 1739 ATAACAACACCTTGATAGCAAAAGAAATTCCTACACAGCTTGCTT 1788
 650 alMetAsnHisIleAsnAsnGlnIaArgMetAsnGlyPheAsnGlyTyrPhe 666
 1789 GCGGCAAGATACGACCAAAAGC...AACGGCGGCTCAACTTGCTTT 1835
 667 GlyGluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrPhe 683
 1836 CCAGCCCGCCAGAGACGCAACCCCTGCTTCGCGGCAACAATT 1885
 683 eaSnGlyLysSerAspGlnAsnArgPheLeuThrGlyGlyThrAsnL 700
 1886 TAAACGGCAATCAGCAAAACAGCGCAACTGTTTTCACGCGCAGA 1935
 700 euAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyArg 716
 1936 CCAACACCGCAGCGCTCAATTCATTTAAAGCACCATTGGTCGCAAAAGA 1985
 717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
 1986 GGGC.....ATTCTCGCGGGGAATCTGTGGGACACAGCTGGATCA 2029
 733 pProHisPheThrGlnAsnAsnGlnValIaIaGluAspAspTrpIleA 750
 2030 ACCGCAATTTAAAGCGGAAACTTCCAATTTAAAGCGGAGAGGGGGT 2079
 750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeu 766
 2080 GTTTC...CGCAATGTTGCCAAAGTGAAGGCGATGGCATTTGAGCAA 2126
 767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
 2127 TCAGCCCAAGCAGTGTGTTGGTGCACCGCATCAAGCCACACAATCT 2176
 783 nasAlaGlnValHisIleGlyTyr...LysThrGlyAspThrValC 798
 2177 GTACAGCTGGAGTGGAGGGTGTGACAAATGTCGAAAAACCATT 2226
 798 ysvAlaArgSerAspTyrThrGlyTyrValThrCysHisAsnSerAsnLeu 814

2227 ACCGCAATTAAGTATGCTTCATTCGACTAAGACCGCATACCGGCA 2276
 815 Ser...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgLys 830
 2277 TGTGATCTTGCCGATCAGCTCATTTAAATCTCACAGGCTTCCACAC 2326
 830 nValAsnLeuThrGlnAsnAlaSerPheThrLeu...GlyLysAlaAsnL 846
 2327 TCACGCGCAATCTAGTCAAAATGGCGATACAGCTTATACGTCAGCAC 2376
 846 eupPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGlu 862
 2377 AACGCC.....ACCCAAACGCGCAACCTTAGCCTCGTGGGC.. 2412
 863 AsnSerHisThrPHisLeuThrGlyAsnSerAsnValAsnGlnLeuAsnLe 879
 2413ATGCCCA.....G 2422
 879 uThrAsnGlyHisIleHisLeuAsnAlaGlnAsnAspAlaAsnLysValT 896
 2423 CAACATTTAATCAACGCCATTTAAACGCAACATCGGCTTCGGGCAT 2472
 896 hrThrTyrAsnThrLeuThrValAsn...SerLeuSerGlyAsn 909
 2473 GCTTCATTT.....ATCTAACGCGACACGC 2498
 910 GlySerPheTyrTyrTrpValAspPheThrAsnAsnLysSerAsnLysVa 926
 2499 CGTACAAAC.....GCGAGTCTGACGCTT..... 2523
 926 lValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspL 943
 2524TCGGGC 2529
 943 ySThrGlyGluProAsnHisAsnGluLeuThrLeuPheAspAlaSerAsn 959
 2530 AACGCTAAGGCAACGTAAGCATTCGCACTCAACGGTAATGTCCTCT 2579
 960 AlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerVal.... 974
 2580 AGCGATTAAGCAGTATTCATTTGAAAGCAAGCGGCTTACCGGACAAA 2629
 975 ...AspArgGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgT 990
 2630 TCAGCGCGCGCAGAGATACGCACTTAAAGACAGCAGATGAGC 2679
 990 YrAsp.....LeuTyrAsnProGlu..... 996
 2680 CTGCGCTAGCGCAGGAATTAGCATTAAACCTTGACAAAGCCACCAT 2729
 997ValGluLysArgAsnGlnThrValAspThrThrAsnI 1009
 2730' TACACTCAATTCGCGCTATCGGCACAGATGGCGGCAACCGGCA 2779
 1009 eThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnSerAsnA 1026
 2780 GT..... 2781
 1026 snGluGluIleAlaLysValGluThrProValProProProAlaProAla 1042
 2782GCAGACATGCGCGCGCGCGCGCTTCGGCGCGCTC 2816
 1043 ThrGluSerAlaIleAlaSerGluGlnProGluThrArgProAlaGluTh 1059
 2817 GCGCGCTTCCTATATACGCTTACA..... 2841
 1059 rAlaGlnProAlaMetGluGluThrAsnThrAlaAsnSerThrGluThrA 1076
 2841 2841
 1076 laProLysSerAspThrAlaThrGlnThrGluAsnProAsnSerGluSer 1092

2842CCGCCAACTTCGGT 2855
1093 ValProSerGluThrThrGluValAlaGluAsnProProGluGluAs 1109
2856 AGAATCC.....CGTTCAACACGCTGACGGCTAA 2884
1109 nGluThrValAlaLysAsnGluGluAlaThrGluProThrProGluA 1126
2885 ACGCAAAATTG.....AAGCGTCAG 2904
1126 sncGlyGluValAlaLysGluAspGluProThrValGluAlaAsnThrGln 1142
2905 GGAACATTCGCGTTATGTCGCACTTCGCGCTACCGCAGCAGCAAAAT 2954
1143 ThrAsnGluAlaThrGlnSerGlu.....GlyLysThrGluGluThr 1156
2955 GAAGCTGGCGGAAAGT...TCGGAAGCACTTACACCTGGCGGTCACA 3001
1156 rGlnThrAlaGluThrLysSerGluProThrGluSerValThrValSerG 1173
3002 ATACCGCAACGACCT.....GCAAGCTCGAACAATG 3036
1173 Lu.....AsnGlnProGluLysThrValSerGlnSerThrGluAspLys 1187
3037 ACGGTAGTGAAGAAAGAC..... 3057
1188 ValValValGluLysGluGluLysAlaLysValGluThrGluGluThrGln 1204
3057 3057
1204 nLysAlaProGluValThrSerLysGluProLysGlnAlaGluProA 1221
3057 3057
1221 LaProGluGluValProThrAspThrAsnAlaGluGluAlaGluAlaLeu 1237
3057 3057
1238 GlnGlnThrGlnProThrThrValAlaAlaAlaGluThrThrSerProAs 1254
3058 AACAAACCGGTGTCGAAACCTTAATTTCACCTGCAAAACGAC... 3103
1254 nSerLysProAlaGluGluThrGlnGlnProSerGluLysThrAsnAla 1270
3104ACGTGATGCCGCGCGTGTACCACTATCGCGCAAA 3144
1271 GluProValThrProValValSerGluAsnThrAlaThrGlnProThrGln 1287
3145 GACGGCGAGTTCGCGCATTAATCCGGTCAAGAACAGAGCTTTCGCA 3194
1287 uThrGluGluThrAlaLysValGluLysGluLysThrGlnGluValProG 1304
3195 CAACCTCGGCAAGGAGCAAGCAAAACAGCGGCAAAACACACACGCC 3244
1304 InValAlaSerGlnGluSerProLysGlnGluProAlaAlaLysPro 1320
3245 AAAGCTTGACGCGTATGCGCGCGCGCATGCCGTCGAAAG... 3291
1321 GlnAlaGlnThr.....LysProGlnAlaGluProAlaAlaGluAs 1334
3292ACAGAACGTTGCCGCAACGCGCGCGCAGCA... 3324
1334 nValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnProG 1351
3324 3324
1351 InThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsnSer 1367
3325GG 3326
1368 LysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAl 1384
3327 CGGGGAAATGTCGCAATTATGCAAGCGGAGGAA..... 3360

1384 aArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnThrS 1401
3360 3360
1401 eAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsnVal 1417
3360 3360
1418 GluGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThrMe 1434
3360 3360
1434 tThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValThrG 1451
3360 3360
1451 LuAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValAla 1467
3360 3360
1468 AsnAsnSerGluSerSerGluSerLysSerArgArgArgSerValSe 1484
3361 GAGAAAAACGGGTGCGCGCATTAAGACACCGCTTCGCGAAACAGC 3409
1484 rGlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrGlnG 1501
3410 GCGAAGCGGAACCCGCGCGCTACCAACGCTCCCGCGCGCGCGC 3459
1501 LuThrThrValAspAsnSerValSerThrProLysProArgSerArgArg 1517
3460 GCCCGCGGATTTGCCGCACTGCAACCCCAACCCGACGCCCAACCGCA 3509
1518 ThrArgArgSerValGlnThrAsnSerTyrcGluProValGluLeuProth 1534
3510 G..... 3510
1534 rGluAsnAlaGluLysAlaGluLysAlaGlnSerGlyAsnAsnValAlaA 1551
3511CGGACCTGATCAGCCGTTATGCCATTAAGCGGT 3543
1551 snSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaVal 1567
3544 TTGAGTGAATTTCCGCGC.....ACGCTCAACGCGCTTT 3578
1568 IleSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLys 1584
3579 CGCGGTACAGAGCAATTAAGCCGCTA...TTTGCGGAAGACCGCGCA 3625
1584 sAlaValSerGlnHisIleSerGlnLeuGluMetAsnGluGlyGlnT 1601
3626 ACCCGCTTGGACAAGCGGATCCGGGACACCAACACATACGCTTGCA 3675
1601 yAsnValTrpIleSerAsnThrSerMetAsnLysAsnTyrcSerSerGlu 1617
3676 GATTCGCGCGCTACCGCCCAACCAACCGAAGCTTCGCAAAATGGTATGA 3725
1618 GlnTyArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTrpAs 1634
3726 GAAAAACCTGCGCAGCGG...CGCGTCGCAATCCGTTTTCGCAACCC 3772
1634 rGlnThrIleSerAsnAsnValGlnLeuGlyGluValPheThrTyValA 1651
3773 GAGCGGAAAAACCTTCGACGACGCGCATGCGCACTCGCGACGCTTGC 3822
1651 rGAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThr...LeuAla 1666
3823 CACGCGCGCGTTTCGGAATACGCGCATCGAC...AGGTTCTACATCGG 3869
1667 GlnValAsnPheTyrcSerLysTyrcTyrcAlaAspAsnHisTrpTyrcLeuG 1683
3870 CATCAGCGCGCGCGGCTTTAGCAGCGGCAAGCTTTCAGACGCGCATCG 3919

```

1683 ylleasrleuGlytyrGlyLyspheGlnSerAsnLeuGlnThrAsnAsna 1700
3920 GAGGCAAAATCCCGCCCGCTCTGCTATAGCGCATTCAGCAGCATAC 3969
      :::::::::::::::::::: ::::::::::::::::::::
1700 snAlaLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLys 1716
3970 CGCGCGCGCTTCGGGATTCGGCATCGAAGCGCATCAATCGGCGCA 4019
      :::::::::::::::::::: ::::::::::::::::::::
1717 AlapheAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTy 1733
4020 CTATTTCGTCCAAAAGCGGATTCAGCTACGAAACGTCATATCGCA 4069
      :::::::::::::::::::: ::::::::::::::::::::
1733 rSerTyLysSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysV 1750
4070 CCCCCGCGCTGCATTCACCGCTACCGCGCGCATTAAGGAGTTAT 4119
      :::::::::::::::::::: ::::::::::::::::::::
1750 AlaSnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTy 1766
4120 TCATTCAAACCGGCGCAACACATTTCCATCAGCTTATTGAGCTGTC 4169
      :::::::::::::::::::: ::::::::::::::::::::
1767 ThrTyrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaAr 1782
4170 CTATACCGATGCGCTTCGGGCAAGTCGACACAGCGTCATATACGCG 4219
      :::::::::::::::::::: ::::::::::::::::::::
1782 gTyL...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTy 1798
4220 TATTGGCTGAGATTCGGCAAAACCGGACGTCGGGATGGGCGTAAAC 4269
      :::::::::::::::::::: ::::::::::::::::::::
1798 sPheAlaTyLysAsnValGluAsnGlnGlnGlnIleThrAsnAlaGlyLeuLys 1814
4270 GCCGAAATCAAGGTTTCACGCTGTCCTCCACGCTGCGCGCGCAAAAG 4319
      :::::::::::::::::::: ::::::::::::::::::::
1815 LeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAl 1831
4320 CCGGCACTGGAAAGCGCAACACAGCGGCGCATCAATTAGGCTAC 4365
      :::::::::::::::::::: ::::::::::::::::::::
1831 ALysGlnAlaGlnGlyGlnIleThrAlaGlnValLysLeuSerPhe 1846

```

```

seq_name: /cgn2_6/ptodata/1/1aa/PCITUS_COMB.pep:PCT-US95-10661A-6

```

```

seq_documentation_block:

```

```

; Sequence 6, Application PC/TUS9510661A

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: Washington University, et al.

```

```

; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

```

```

; NUMBER OF SEQUENCES: 9

```

```

; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Flehr, Holbach, Test, Albritton & Herbert

```

```

; STREET: 4 Embarcadero Center, Suite 3400

```

```

; CITY: San Francisco

```

```

; STATE: California

```

```

; COUNTRY: United States

```

```

; ZIP: 94111-4187

```

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk

```

```

; OPERATING SYSTEM: IBM PC compatible

```

```

; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: PCT/US95/10661A

```

```

; FILING DATE: 16-AUG-1995

```

```

; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/296,791

```

```

; FILING DATE: 25-AUG-1994

```

```

; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Treacatlin, Richard F.

```

```

; REGISTRATION NUMBER: 31,801

```

```

; REFERENCE/DOCKET NUMBER: FP-59941/RFT

```

```

; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (415) 781-1989

```

```

; TELEFAX: (415) 398-3249

```

```

; TELLEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-6

```

```

alignment_scores:
  Quality: 1486.00      Length: 1916
  Ratio: 1.490
  Percent Similarity: 52.035      Percent Identity: 24.843

```

```

alignment_block:
US-09-303-518D-649 x PCT-US95-10661A-6

```

```

Align seg 1/1 to: PCT-US95-10661A-6 from: 1 to: 1848

```

```

64 CGCTTCGCGCTGCTATGACCATATGCGCTGCTTCGCAATTCCTCC 113
      :::::::::::::::::::: ::::::::::::::::::::
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyAlaLeuThrPr 21
114 CMAAGCTGGGCGGACACACTTATTCGCGATCACTACCAATACTATC 163
      :::::::::::::::::::: ::::::::::::::::::::
21 oTyThrGluAlaIleAlaLeuValArgAspAspValAspTyGlnIlePhe 38
164 GCGATTTTCCGAAATATGAAGCAAGTTTCAGTCGCGGCGCAAGATAT 213
      :::::::::::::::::::: ::::::::::::::::::::
38 ArgSPheAlaGlnAsnLysGlyLysPheSerValGlyAlaIleThrVal 54
214 GAGGTTTACACAAAAGGGGAGTGTGCGCAATATCATGACAAAAGC 263
      :::::::::::::::::::: ::::::::::::::::::::
55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71
264 C...CGATGATGATTTTCTGCTGCTG...CGTACGCGCGTGGCGG 307
      :::::::::::::::::::: ::::::::::::::::::::
71 ylleProMetIleAspPheSerValValAspValAspLysArgIleAla 88
308 CATTCGTGGCGGATCAATATGATTCGACGCTGGACAT...AACGGCGG 354
      :::::::::::::::::::: ::::::::::::::::::::
88 hLeuValAsnProGlnTyValValGlyValLysHisValSerAsnGly 104
355 TATAACAACGTTGATTTCGT...GGGGAAGGAAGAAA 389
      :::::::::::::::::::: ::::::::::::::::::::
105 ValSerGlnLeuHisPheGlyAsnLeuAsnLysAsnMetAsnGlnLys 121
390 TCCGATGCAACATCGTTTACT...TATAAATTCG 421
      :::::::::::::::::::: ::::::::::::::::::::
121 nAlaLysSerHisArgAspValSerSerGlnLysAsnArgTyTyThrV 138
422 TGAACGGAATATATAT...AAAGCA 444
      :::::::::::::::::::: ::::::::::::::::::::
138 AlGlnLysAsnAsnPheProThrGluAsnValThrSerPheThrLysGlu 154
445 GGGATAAAGGCCATCCTTATGCGCGCATTCATCATGCGCGCTTGCA 494
      :::::::::::::::::::: ::::::::::::::::::::
155 GluGlnAspAlaGlnLysArgGlnLysArgGlnLysArgGlnLysArg 171
495 TAAATTTGTCACAGATGCAAGACGTTGAAATGACAGTATATATGATG 544
      :::::::::::::::::::: ::::::::::::::::::::
171 PLYSPheValThrGluValAlaProIleGluAlaSerThrAlaAsnAsna 188
545 GCGCG...AAATATATGATCAAAATATATACCTGACGCTGTCGATTC 591
      :::::::::::::::::::: ::::::::::::::::::::
188 snLysGlyLysThrGluAsnAsnSerAspLysTyProAlaPheValArgLeu 204
592 GGGCAGCGCAGCATAT...TG 611
      :::::::::::::::::::: ::::::::::::::::::::
205 GlySerGlyThrGlnPheLysLysLysLysSerArgTyGlnLeuI 221
612 GGGATCTGATGAAGATGAGCCCAATACCGCAAGTTCATATCAT... 657
      :::::::::::::::::::: ::::::::::::::::::::

```


221 eleuthrhlulysasplysglmglyAsnleuleuArasntrpaspValG 238
658ATTGCAAGTGGCTATTCTTGGCTGCTGGT 687
238 lylGlyAspAsnleuGluleuValGlyAsnAlatYrThrlyrGlylleAla 254
688 GGCATTAACCTTTGCACAAATGATGACAGTGGTGGACAGTCACATTAAG 737
255 GlyThrProTyrLysValAlaAsnHISgluAsnsglyLeuIleGlyPheG1 271
738 TAGTGAATAAATTAACATAGC...CCATATGCTTTTACCACAGCA 783
271 yAsnserLysgluGluHISserAspProLysGlyIleLeuSerGlnAsp 288
784GGTCATTTGGCGACAGTGGCTCACCAATGTTTTC 819
288 roleuthrAsntrYrAlaValleuGlyAspSerGlySerProleuHeVal 304
820 TATGATGCCCAAAAGCAAAAGCTGTAATTAATGCGGTATTCACAAACGG 869
305 TyrAspArgGlyLysGlyLysThrLeuPheLeuGlySerTyrAspPheTr 321
870 CAACCCCTATATAGCAAAAGCAATGCTTCAGCTGCTCGTAACAT 919
321 PALaglyLysAsnLysLysSerTrpGlnLutrpAsnIleTyrLysHISG 338
920 GATTCTATGATGAATCTTGTGCTGACATCCATTCATTCATTCACAA 969
338 lUpheAlaglyLysIleTyr.....GlnGlnTyrSerAlaglySerleu 352
970 CCAGCTCAAAATGGGAATACTCTTTTACAGCAGATAATATGACACAG 1019
353 lIleGlySerAsnThrGlnTyrThrTrp.....GlnAlaThrG1 365
1020 AAAATTCATATGCAACATGAACACATTCCTGCTAATAGATTAATAA 1069
365 y.....SerT 367
1070 CACGAAACCGTT.....CAATGTTTAAATGTTTCTTATTCAGAG 1107
367 hrSerThrIleThrGlyLysGlyLysProLeuSerValAspLeuThrAsp 383
1108 ACAGCAAGAAGACCTGTTATCATGCTCAGAGTGTCAACAGTTATCG 1157
384 GlyLysAspLysPro..... 388
1158 ACCCAGACTGAATATGAGAAAATATTCTTATTATTCAGCAGAGAAAG 1207
389AsnHISglYysSerIleThrLeu.....LysGlySerG 400
1208 CGCAATTGATCTTACAGCAACATCAATCAAGTGTGAGAGATTATAT 1257
400 lYThrLeuThrLeuAsnHISleAspGlnGlyAlaGlyLysLeuPhe 416
1258 TTCAGAGAGATTTTACGGTC...TCGGCTGAATAATACGAACTTGGCA 1304
417 PheGluGlyAspTyrGluValLysGlyThrSerAspSerThrTrpLys 433
1305 AGGCGCGGCGTTCAATACAGTGAAGACAGTACCTGATTGAAGATTA 1354
433 sglYAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450
1355 ACAGCGGTGGCAACGACCGCTGTCCAAATGCGCAAGCAGCTGCAC 1404
450 lAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466
1405 GTTCAAGCGCAAGGGAACCAAGCGTCATCAGCGTGGCGGCGGCTAC 1454
467 ValGluGlyLysGlyLysAsnGluGlyLeuLeuLysValGlyAspGlyTh 483
1455 AGTCATTTGGATCAGCAGCAGCAGATTAAGCAAAAGCAAGCCTTTA 1504
483 rValIleLeuLysGlnLysAlaAspAlaAsnAlaValGlnAlaPheS 500
1505 GTGAATCGCGCTTGTCAGCGCAGGCTACGGTCACTGAATGCCGAT 1554
500 erGlnValGlyIleValSerGlyArgSerThrLeuValleuAsnAspArg 516
1555 AATCAGTTCAACCCGACAAACCTATTTGGCGTTTCGGCGGAGGATT 1604
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyLysArgle 533
1605 GGATTTTAAAGGGGCGATTCGCTTTCGTCACCGCTATTCAAATACGATG 1654
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHISleArgAsnIleAspArg 550
1655 AAGGGCGGATGATTTGTACACCAATCAAGCAAAAGATACACCGTTAC 1704
550 spGlyAlaArgValAlaAsnHISAsnMetThrAsnThrSerAsnIleThr 566
1705 ATTACAGGC..... 1713
567 lIleThrGlyLysSerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1713 1713
583 nIleGluAlaGlnAspAspAspHISProLeuArgIleArgSerIlePro 600
1713 1713
600 yArArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrThrLeuLys 616
1713 1713
617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyLysLysAs 633
1714AATTAAGATATTGCTACACCGCGCA 1738
633 ngLAsnTrpLeuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1739 ATTAACACAGCTTGATAGCAAAAGAAATGCTTACAGACGGTGGTT 1788
650 alMetAsnHISleAsnAsnGlnArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GCGCAGAAAGATACGACCAAAAGC...AACGGCGGCTCAACCTGTTTA 1835
667 GlyGluGlnGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrph 683
1836 CCAAGCCCGCGCAGAAAGCCGACCTGCTTCCGGCGGCAACAAT 1885
683 eaSnglyLysSerAspGlnAsnArgPheLeuThrGlyGlyThrAsnL 700
1886 TAAACGGCAACATCAGCAAAACAAAGCGCAACTGTTTTCAGCGGCGA 1935
700 euAsnGlyAspLeuAsnValGlyLysGlyThrLeuPheLeuSerGlyArg 716
1936 CCAACACCGCAGCGCTACATTCATTTAAACGACCATTTGTCGCAAAAGA 1985
717 ProThrProHISAlaArgAspIleAlaGlyIleSerSerThrLysLys 733
1986 GGGC.....ATTCTCGGGGGAATCGTGGCGCAACGCTGATCA 2029
733 pProHISpHeThrGluAsnAsnGluValAlaValGlnAspAspTrpIleA 750
2030 ACCGCACTTTAAAGCGGAAACTTCCAATTTAAAGCGGAGCGAGCGGTG 2079
750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnHISAsnLeu 766
2080 GTTTC...CGCATTTGCCAAAGTGAAGGCGATGGCATTTGACAA 2126
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCAGCGCCAGCAGGTTTGGTGGTGGCACCAGCATCAAGCCACACACACT 2176
783 naSnAlaGlnValHISIleGlyTyr.....LysThrGlyAspThrValC 798

1368 LysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrLysProAla 1384
3327 CGGGAAATATCGCATATGACGGAGGAA..... 3360
1384 aArgLysAsnValSerThrValAsnThrLysGlnProGlnSerGlnThr 1401
3360 3360
1401 eAlaThrValSerThrGlnProAlaLysGlnThrSerSerAsnVal 1417
3360 3360
1418 GlnGlnProAlaProGlnAsnSerIleAsnThrGlySerAlaThrThrme 1434
3360 3360
1434 tThrGlnThrAlaGlnLysSerAspLysProGlnMetGlnThrValThrG 1451
3360 3360
1451 LuAsnAspArgGlnProGlnAlaAsnThrValAlaAspAsnSerValAla 1467
3360 3360
1468 AsnAsnSerGlnSerSerGlnSerLysSerArgArgArgSerValSe 1484
3361 .GAGAAAAACGGGTGACGGGATTAAGACACGGCTTGGCGAAACAGC 3409
1484 rGlnProLysGlnThrSerAlaGlnGlnThrThrValAlaSerThrGlnG 1501
3410 GCGAAGCGGAAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGC 3459
1501 lThrThrValAspAsnSerValSerThrProLysProArgSerArgArg 1517
3460 GCCCGCGCGGATTTCCCGCAACTGCACCCGACCGCGCGCGCGCA 3509
1518 ThrArgArgSerValGlnThrAsnSerArgLysProValGlnLeuProTh 1534
3510 G..... 3510
1534 rGlnAsnAlaGlnLysAsnAlaGlnLysGlnSerGlyAsnAsnValAla 1551
3511CGCGACCTGATCAGCCGTTATGCCAATAGCGGT 3543
1551 snSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaVal 1567
3544 TTGACTGATTTTCGCC.....ACGCTCACACGCGTTT 3578
1568 IleSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLys 1584
3579 CGCCGTACAGAGCAATTAGACCGGTA...TTGGCGAAGACCGCGCA 3625
1584 sAlaValSerGlnHisIleSerGlnLeuGlnMetAsnAsnGlnGlnGlnT 1601
3626 ACGCGTTTGACAACGGCGCATCCGGGACACCAACTACCGTTGCGAA 3675
1601 rAsnValAlaIleSerAsnThrSerMetAsnLysAsnTyrSerSerGln 1617
3676 GATTTCCGCGCTACCGCCACAACACGCTGCGCAATCGGTATGCA 3725
1618 GlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTyrPps 1634
3726 GAAAAACCTGGAGCGG...CGCGTCGCGCATCTGTTTTCGACACAC 3772
1634 pGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrThrVal 1651
3773 GGACCGAAAAACCTTGACGACGCGCATCGCAACTGCGACGCGTGGC 3822
1651 rGAsnSerAsnAsnPheAspLysAlaSerLysAsnThr...LeuAla 1666
3823 CACGCGCGCTTTTCGGGCAATACGCGATGAC...AGGTTCTACATCGG 3869
1667 GlnValAsnPheTyrSerLysTyrThrAlaAspAsnHisTyrPyrLeuG 1683

3870 CATTACGCGCGCGCGGCTTTTAGACACCGCGACGCTTTAGACGCGATCG 3919
1683 yLleAspLeuGlyTyrGlnLysPheGlnSerAsnLeuGlnThrAsnAsn 1700
3920 GAGCGAAATATCGCGCGCGCTGCGATTAACGCGATTAACGCGCGATAC 3969
1700 snAlaLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLys 1716
3970 CGCGCGCGTTTCGGCGATTCGCGCATGCAACCGCACATCGCGCGCGCG 4019
1717 AlaPheAsnLeuGlnAsnPheAlaValLysProThrValGlyValArg 1733
4020 CTATTGCTCCAAAAAGCGGATTAACCGCTTCGAAACGTAATATGCGCCA 4069
1733 rSerTyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLys 1750
4070 CCCCCGCGCTTGCAATTCACCGCTACCGCGCGCGCATTAACGCGATAT 4119
1750 aAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyr 1766
4120 TCATTCAACCGCGCGCACACATTTCCATCGCGCTTATTAGCGCTGTC 4169
1767 ThrTyrHisLeuGlyLys...PheSerIleThrProIleLeuSerAla 1782
4170 CTATTACCGATCGCGCTTCGCGCAAGTCGCAACACGCGTCAATACGCGG 4219
1782 gTyr...AspAlaAsnGlnGlnLysGlnLysIleAsnValSerValTyr 1798
4220 TATTGCTCAGATTTCGCGCAACCGCGAGTCGCGAATGCGCGCTTAAC 4269
1798 spPheAlaTyrAsnValGlnLysGlnGlnGlnThrAsnAlaGlyLys 1814
4270 GCGGAATCAAGGTTTCAGCGTGTCCCTTCAGCGTGGCGCGCGCAAG 4319
1815 LeuLysTyrHisAsnValLysLysLeuSerLeuIleGlyLeuThrLysAl 1831
4320 CCGCAACTGGAAGCGCAACACACGCGCGCGCATTAATGAGCTAC 4365
1831 aLysGlnAlaGlnLysGlnLysThrAlaGlnValLysLeuSerPhe 1846
seq_name: /cgn2.6/ptodata/1/aa/6B-COMB.pep:US-08-460-269C-4
seq_documentation_block:
; Sequence 4, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333

TELEFAX: (703) 243-6410
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 911 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-08-460-269C-4

alignment_scores:
 Quality: 295.50 Length: 1079
 Ratio: 0.612 Gaps: 50
 Percent Similarity: 44.764 Percent Identity: 21.965

alignment_block:
 US-09-303-518D-649 x US-08-460-269C-4 ..

Align seg 1/1 to: US-08-460-269C-4 from: 1 to: 911..

```

1417 GGGGAAACAGAGCTGCTGACGCTGGCGAGCGTACATCATTTTGA 1466
|||||:|||||:|||||:
45 GLYGLNARGLNHISGLYILEHISLE..... 53
1467 TCACGACGACAGACGATTAAGCAAAAAACAGCTTTAGTGAATCGGCT 1516
:|||||:|||||:|||||:
54 LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIle 70
1517 TGTGTCAGCGGACGGGTACGCTGCACTGCAATGCCATATCATGTTCAAC 1366
|||||:|||||:|||||:
70 yValSerGlyArgLInlaGlnGlyValLeuLeuGluasn..... 83
1567 CCCGACAACTCTATTTCGCTTCGCGGCGAGCGTTGGATTAAACGG 1616
:::|||||:|||||:|||||:
84 ...ProLalaGluLeuArgPheGlnAsnGlySerValThrSerSerI 99
1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGCGCATGA 1666
|||:|||||:|||||:
99 y.....GlnLeuPheAspGluGlyVal..... 106
1667 TTTCACACACATCAAGACAAAGATCCACCGTTACCATTTACAGCAAT 1716
:::|||||:|||||:|||||:
107 .....ArgArgPheLeuGlyThrValThrValLysAlaGly 118
1717 AAA.....GATTGTGTACACCGGAGATTAACAAGCTTGGATAG 1757
|||:|||||:|||||:|||||:
119 LysLeuValAlaAspHisAlaThrLeuAlaasn..... 129
1758 CAAAAAAGAAATTCCTACAAACGGTTGGTTGGCGAGAAAGATACGACCA 1807
:::|||||:|||||:
130 .....ValSerAspThrArg 135
1808 AAGCAACGGCGGCTCAACCTGTT.....TACACCGCGCGC 1845
:|||||:|||||:|||||:
135 spAspAspGlyIleAlaLeuTyValAlaGlyGluGlnAlaGlnAlaSer 151
1846 GCAGAGACCCACCTGCTGCTTCCGCGGGAACAATTTA.....AA 1889
|||:|||||:|||||:|||||:
152 IleAlaSerSerThrLeuGlnGlyAlaGlyValArgValaGluArgGly 168
1890 CGGCAATCTCCGCAACAAACGCAACTGTTTTCAGCGGC..... 1932
:|||||:|||||:|||||:
168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyLeuHisI 185
1933 .....AGACCAACACCGCAGCGCTACATCATTTAAAGACCAT 1971
:|||||:|||||:
185 legGlyThrLeuGlnProLeu..... 191
1972 TGTGTCGCAAAAAAGAGGATCTCTCGCGGGAATACTGTGGGCAACGA 2021
|||:|||||:|||||:|||||:
192 .....GlnProGluAspLeuProProSerArgValaValLeuGlyAspTh 206

```

```

2022 CTGATCAAC.....CGACAT 2038
:::|||||:
206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValSerVal 223
2039 TTAAGCCGAAACTTCCAAATTAAGCGGACAGCGCGTGTCCCGC 2088
|||:|||||:|||||:|||||:
223 heGlyAlaAsnGlnLeuThrValAspGlyGlnHisIleThrGlyGlyArg 239
2089 AATGTTCCCAAGTGAAGCGGATTTGCATTTGAGCATTCAGCATCCGCGCAAGC 2138
:::|||||:
240 AlaAlaGly.....ValAlaAla 245
2139 AGTTTGTGTGTCGACCGCATCAAGCCACACATCTGTACAGCTTCG 2188
:::|||||:|||||:|||||:
245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGly 262
2189 AC..... 2190
|||
262 spAlaProAlaGlyAlaValaProGlyGlyAlaValaProGlyGlyPhe 278
2191 .....TGCAGCGGTCTGACAAATTTGTGCAAAAAAC 2222
|||||:|||||:
279 GlyProLeuLeuAspGlyTryPtyGlyVal..... 288
2223 CATTCACGACGATTAAGTGAATGCTTCATTGACTAGACCGACATCAGC. 2271
|||||:|||||:
289 .....AspValSer 292
2272 ..GGCAATGTGATCTGCGATCAGCTCATTTAATCTCAGAGGCTT 2319
:|||||:|||||:|||||:
292 spSerThrValAspLeuAlaGln.....SerIleValGlnAla 304
2320 GCCACATCAACGGCAATCTTAGTCAAAATGGCATACAGTTATACAGT 2369
|||||:|||||:|||||:
305 ProGlnLeuGlyAlaAlaIleArgAlaGlyArgGlyAlaArgValaThrVa 321
2370 CAGCCACACAC.....GCCACCCAAACGGCAACCTTAGCTCGGGGCA 2413
|||:|||||:|||||:
321 IserGlyGlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyG 338
2414 ATGCCCAAGCAACATTTAAT.....CAAGCCACATTA 2445
:::|||||:|||||:
338 yGlyAlaArgArgPheProProProAlaSerProLeuSerIleThrLeu 354
2446 AACGGCACACATCGGCTTCGGGCAATGCT..TCATTTAATCTAAGCA 2492
:|||||:|||||:|||||:
355 GlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLeuTyArgValLeuPr 371
2493 CCACGCCGTACAAAAGCGCATCTGACGCTTCGCGACAGGTAAAGGCA 2542
:::|||||:|||||:
371 OGluProValLys.....LeuThrLeuAlaGlyValaGlnGly. 384
2543 ACGTAAGCCATTCCGACTCAACGGTAATGTCTCCCTAGCCGATTAAGCA 2592
:|||||:|||||:
385 .....GlnGlyAspIleValaIleThr..... 391
2593 GATTTCATTTTGAAGACAGCGCTTTACCGGACAAATCAGCGGGGCA 2642
|||:|||||:|||||:
392 .....GluLeuProProIleProGlyAlaSerSerGlyProLe 404
2643 GGTATGCGATTACACTTAAAGACAGCAAGATGACGCTGCGTCAGCA 2692
|||||:|||||:|||||:
404 uAspValaIleLeu...AlaSerGlnAlaArgTyrThr...GlyAlaThr 419
2693 CGGAATTAGCAATTTAAACCTTGACAAACGACCATTAACCTCAATTC 2742
:::|||||:|||||:|||||:
419 rGlyAlaValAspSerLeuSerIleAspAsnAlaThrTyrAlaMet..... 433
2743 GCCTATTCGCACAGATCGGACAGGCGCCAAACCGGACGATGCGACAGATGC 2792
|||:|||||:
434 .....ThrAsp... 435
2793 GCGCGCGCGCGCTGCGCGCTTCGCGCGCTTCCTATATCCGTTACAC 2842

```

```

436 .....:||||: ||| :|||:
2843 CGCCAACTTCGGTAGA.....AsnSerAsnValGlyAlaLeuArgLeuAlaAs 446
|||||:|||||:
446 erAspGlySerValAspPheGlnGlnProAlaGlnAlaGlyArgPheLys 462
2872 AGCTGACGCGTAAAGCGCAATTTGAACGCTGACGAACTATCCGTTTAT 2921
|||||:|||||:
463 CysLeuMetValAsp...ThreuAlaGlySerGlyLeuPheArgMetAs 478
2922 GTCCGAACTCTTCGGCTACCGGACGACAAATTTGAAGCTGGCGAAAGTT 2971
|||||:|||||:
478 nValPheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspA 495
2972 CGGAAGCGCTTACACCTTGGCGGTCAACAAATACCGGACGACAACTGCA 3021
|||||:|||||:
495 LAspGlyGlnHisArgLeuValArgAsnSerGlySerGlnProAla 511
3022 AGCTCGAACAATTTAGCGGTAGTGAAGAAAGACAAACCGCTGTC 3071
|||||:|||||:
512 SerGlyAsnThrMetLeuLeuValGln.....ThrProArgG1 524
3072 CGAAACCTTAATTTACCTTCGCAAAAGCA.....CACTGATGCGC 3115
|||||:|||||:
524 ySerAlaAlaThrPheThrLeuAlaAsnLysAspGlyLysValAspIleG 541
3116 GCGCGGCGCTTACCACTCCGCAAAAGAGCGGAGTTCGCTGCAT 3165
|||||:|||||:
541 yThrTyArgTyrArgLeu..... 547
3166 AATCGGTCMAAGAACAGAGCTTTCGCAAACTCGGCAAGCGAGAAC 3215
547 ..... 547
3216 CAAAAACAGCGGCAAAAGACAGCCCAAGCTTACCGCGCTGATG 3265
|||||:|||||:
548 .....AlaAlaAsnGlyAsnGlyGlnTrp.....SerLeuValG 559
3266 GCGCGGCGCGATGCGCTCGAAAGACAGAAAGCTTGGCAACCGGCGC 3315
|||||:|||||:
559 LysAlaLysAlaProProAlaPro.....LysP 568
3316 CGGCAAGGAGCGGCGGAAATGTGCGATTATGACGGGAGCAAGAACAA 3365
|||||:|||||:
568 AlaProGlnProGlyPro..... 574
3366 AAACGGGTGCAGCGGATTAAGACAGCGCTTGGCGAAACAGCGCGAAG 3415
574 ..... 574
3416 CGAAACCGCGGCGCTACACAGCGCTTCCCGCGCGCGCGCGCGC 3465
|||||:|||||:
575 ...GlnProGlyProGlnProProGlnLProProGlnProProGlnProP 590
3466 GCGGATTTGCGCAACTGCAACCCCAACCGGACCGGCGCGGCGGA 3515
|||||:|||||:
590 rGlnArgGlnProGlnAlaLProAlaProGlnProProAlaGlyArgG1 606
3516 CCGTACAGCGCTTATGCCAATAGCGGTTTGAGTAATTTCCGCCACGC 3565
|||||:|||||:
606 uLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGlyGlyValGlyL 622
3566 TCACACAGGTT.....TTGCGCGTACAGAGCAATTTAGACGCGGTAATT 3609
|||||:|||||:
622 euAlaSerThrLeuTyrTrpAlaGlnSerAsnAlaLeuSerLysArgLeu 638
3610 GCCGAGACCGCGCGCAAC.....GCCGTTTGGACAGCGGCAT 3647
|||||:|||||:
639 GlyLeuArgLeuAsnProAspAlaGlyValArgGlyArgGlyP 655
3648 CCGGAGACCAACACACTACGCTTCGCAAGATTTCGCGCGCGGCGCAAC 3697
|||||:|||||:
655 eAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArgArgPheAspGln 672
3698 AAACGACCTCGCGCAAAATCGTATGCAGAAACCTC..... 3735
|||||:|||||:
672 yValAlaGlyPheGlnLeuGlyLysAsnHisAlaValAlaValAlaGly 688
3736 GCGAGCGGCGCGCGCATCTGTTTCGACACAGCGGACCGCAAAAC 3785
|||||:|||||:
689 GlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgGlyAspArgG1 705
3786 CTTC...GACAGCGCATCGGCACTGCGACGCGCTTGGCGCGCGCGC 3832
|||||:|||||:
705 yPheThrGlyAspGlyGlyGlnHisThrAspSerValHisValGlyGlyT 722
3833 TTTTCGGGCAATACGCGATGACAGGTTTTCATCGGATCGGATCGCGGCG 3882
|||||:|||||:
722 yAlaThrTyrThrAlaAsnSerGlyPheTyrLeuAspAlaThrLeuArg 738
3883 GCGGTTTTCAGCGCGCGCTT.....TCAGACGCG..... 3915
|||||:|||||:
739 LAspArgLeuGlnAsnAspPheLysValAlaGlySerAspGlyTyrAl 755
3916 .ATCGAGCGCAAAATCCGCGCGCGCTGCTGCATTACGGCATTCAGGAC 3964
|||||:|||||:
755 aValLysGlyLysTyrArgThr.....HisGlyValGlyAlaAs 768
3965 GATACCGCGCGCT.....TTCGCGGATTCGCGCATCGAA 3999
|||||:|||||:
768 eLeuGlnAlaGlyArgArgPheAlaHisAlaAspGlyTyrPheLeuGln 784
4000 CGGCACTCGCGCGCAAGCTTATTCGCAAAAGCGGATTCAGGCTTA 4049
|||||:|||||:
785 ProGlnAlaGlyLeuAlaValAlaPheArgValGlyGlySerTyrArgAl 801
4050 CGAAAC...GTCAATATCCGCAACCGCGCGCTTGCATTCAACGCTTAC 4096
|||||:|||||:
801 aAlaAsnGlyLeuArgValArgAspGlnGlySerSerValLeuGly 818
4097 GCGCGGCAATTAAGCGATTTATTCATTCACAGCGGCGCAACATTTCC 4146
|||||:|||||:
818 rGlyLeuGlyLeuGlyValGlyLysArgGlyLeuLeuAlaGlyArgGln 834
4147 ATCAGCGCTTATTTAGCGCTGCTTATACGATCGCGCTGGGCAAGT 4196
|||||:|||||:
835 ValGlnProTyrTrpLysAlaSerValLeuGlnGlnPheAspGlyAlaG1 851
4197 CCGACACGCGGTCAATACCGCGTATTTGCTCAGATTTTCGCAAAACCC 4246
|||||:|||||:
851 yThrValArgThrAsnGlyThrAlaHisArgThrGlnLeuArgGlyThrA 868
4247 GCAGTGCGGAATGGGCGGTAAACCGCGAAATCAAGGTTTCACGCTGTC 4296
|||||:|||||:
868 rG...AlaGlyLeuGly.....LeuGly 874
4297 CTCACGCGTCCGCGCGCAAGGCGCGCAACTGGAAGCG.....CAAA 4340
|||||:|||||:
875 MetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGly 891
4341 CAGCGCGGCGATCAATTTAGGCTTACCGCTGCG 4371
|||||:|||||:
891 rSerLysGlyProLysLeuAlaMetProTrp 901
seq_name: /cgn2_6/ptodata/1/laa/5B_COMP.pep:US-08-728-470-10
seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

```



```
1237 CAAGTGTGAGAGATTATATTTCCAGGAGATTTCAGGTCTCCGCTGA 1286
      |||  ::  ::||| ||| |||
862 IleIyGInGlyValValIyLeuGInGlyAspIle..... 873
1287 AATAACGAACCTTGCAAGCGCGGCTTCATATCAGTAGAACAGTA 1336
      ||| |||  ::||| ||| |||
874 .AsnAsnIyS.....GlyGlyLeuAsnIleThrThrAsnAlas 886
1337 CCGTACTCTGSA.....GTAAACGC...GTGGCAACGACCGCCTG 1377
      :: ||| |||  ::||| ||| |||
886 eGcIyThGInIySthrIleIleAsnGlyAsnIleThrAsnIleu..... 900
1378 TCCAAAATCGCAAGCACGCTG.....CAGTTCAACGCCAAG 1418
      ||| ||| |||  ::||| ||| |||
901 .....LysGlyAspLeuAsnIleIySAsnIleIySAlaAspAl 913
1419 GGAACAAACCA.....GGCTGATCAGCGTGGCGACGGTACATGATT 1462
      ||| |||  ::||| ||| |||
913 eGInIleGInIleGlyAsnIleSergInIySGLuGlyAsnLeuThr 930
1463 TG..... 1464
930 leSerSerAspIySAlaSnIleThrAsnGInIleThrIleIySAlaGly 946
1465 .....GATCAGCAGCGACGATTAAGGCCAATAA 1493
947 ValGluGlyIyArgSerAspSerSergInIaGluAsnAlaAsnLeuTh 963
1494 ACAACCTTTGTGAATCGGCTTG.....GTACGGCGCA 1528
      ||| |||  ::||| ||| |||
963 rIleGInIySthrIySGLuLeuIySLeuAlaGlyAspLeuAsnIleSergIy 980
1529 GGGTACGCTGCACTGAATGCCATAT..... 1557
      ::  ::||| ||| |||
980 heAsnIyAlaGluIleThrAlaIySAsnGlySerAspLeuThrIleGly 996
1558 .....CAGTTCAACCCCGCAAACTCTATTTCGCG...TT 1589
      ::  ::  ::||| ||| |||
997 AsnAlaSerGlyIyAsnAlaAspAlaIySAlaIyThrPheAspIySAla 1013
1590 TCGCGCGGAGCTTGGATTAAACGGGCAATTCGTTCCGTCCACCGTA 1639
      ::||| |||  ::||| ||| |||
1013 IlyAspSerIySleSerThrAspGlyIyAsnValIyThrLeuAsnSerg 1030
1640 TTCAAAATACCGATGAAGGGCGATGATTGTCAACCAATCAAGACAAA 1689
      ::||| |||  ::||| ||| |||
1030 IuValIySThrSerAsnGlySerSer.....AsnAlaGlyAsnAspAsn 1044
1690 GAATCCACCGTTACCATTAACAGGCATTAAGATTTGCTACACCGGCAA 1739
      ::||| |||  ::||| ||| |||
1045 SerThrGlyLeuThrIleSerAla...LysAspValIyThrVal..... 1057
1740 TAACAACAGCTTGGATCAAAAAGAAATGCTTACACAGGTGGTTG 1789
      ||| |||  ::||| ||| |||
1058 .AsnAsnAsnValIyThrSerHisIySThrIleAsnIleSerAlaIaIaG 1074
1790 GCGAAGAAGATACGCAAAACGAGCGGCGCTCAACCTTGTATTACAG 1839
      ||| |||  ::||| ||| |||
1074 IyAsnVal...ThrThrIySGLuGlyThrThrIleAsn..... 1085
1840 CCGCGCGGAGAAAGACCGCTGCTGCTTCGCGCGCAACAAATTMAA 1889
      ||| |||  ::||| ||| |||
1086 ...AlaThrThrGlySerValGluValIyThrAlaGluAsnGlyThrIleIy 1101
1890 CGGCAACATCAGCAACAAACGCAAACTGTTTTCAGCGCGACGACAA 1939
      ::||| ||| |||  ::||| ||| |||
1101 sGlyAsnIleThrSergInAsn.....ValIy 1110
1940 CACCGCAGCGCTACATATTAAAGACCATTTGGTGCACAAAAGAGGC 1989
      ||| |||  ::||| ||| |||
1110 hrValIyThrAlaThrGluAsnLeu..... 1117

1990 ATTCTCGCGGGGAATTCGTGTGGACACAGCATGTCAACCGCACATT 2039
      ::||| |||  ::||| ||| |||
1118 .....ValIyThrThrGluAsnAlaValIleAsnAlaThrSe 1129
2040 TAAACGGAAATCTTCCAAATTAAGCGGCGACAGCGGTGTTCCCGCA 2089
      ::||| |||  ::||| ||| |||
1129 rGlyThrValAsnIleSerThrIySThrGlyAsp.....IleIySGLy 1144
2090 ATGTGCCAAAGTGAAGGCGATTTGAGCAATTCAGCCCAAGCA 2139
      ::||| |||  ::||| ||| |||
1144 IyIleGInSerThrSergIyAsnValAsnIleThrAlaSergIyAsnThr 1160
2140 GTTTTGTGTGTCGACCGCATCAAGCCACACATCTGTACAGTTCCGA 2189
      ::||| |||  ::||| ||| |||
1161 Leu...LysValSerAsnIleThrGlyGluAspValIyThrAlaAs 1176
2190 CTGACCGGCTGACAAATGTGTGCAAAAACCATTCACGAGATMAAG 2239
      ::||| |||  ::||| ||| |||
1176 pAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThrG 1193
2240 TGATTCCTTATTG...ACTAAGAC...GACATCAGCGCAATGTGAT 2283
      ::||| |||  ::||| ||| |||
1193 IyAsnAlaAsnIleThrThrIySThrGlySpleAsnGlyIySValGlu 1209
2284 CTTCGCGAT...CAGCTCATTTAAATCTCAGAGGGCTTCGCCACTCA 2330
      ::||| |||  ::||| ||| |||
1210 SerSerSergIySerValIyThrLeuValAlaIyThrGlyAlaIyThrLeuAlaIy 1226
2331 CGGCAATCTTATGCAAAATGCGCATACA.....CGTT 2362
      ::||| |||  ::||| ||| |||
1226 IyGlyAsnIleSergIyAsnThrValIyThrIleThrAlaAspSergIyLys 1243
2363 ATACAGTACCCACACAGCCACCCCAAAACGCACTTACCTCGTGGCG 2412
      ::||| |||  ::||| ||| |||
1243 euthrSerThrValIySethrIleAsnGlyThrAsnSergValIyThrThr 1259
2413 AATGCCCAAGCAACATTTAATCAAGCCACATTAAGCGCAACATTCGCG 2462
      ::||| |||  ::||| ||| |||
1260 SerSergInSergIyAspIleGluGlyThrIleSergIyAsnThr..... 1274
2463 TTGCGGCAATGCTTCATTATCTAAGCGACACGCGGTCAAAACGGCA 2512
      ::||| |||  ::||| ||| |||
1275 .....ValAsnValIyThr.....AlaSerThrGlyAla 1283
2513 GTCTGACGCTTTCGGGAACGCTAAGGCAAGCTAAGCCATTCGCAATC 2562
      ::||| |||  ::||| ||| |||
1283 spleuthrIleGlyAsnSerAlaIySValGluAlaIySAsnGlyAla... 1298
2563 AACGGTAATGCTCTCCCTAGCCGATTAAGCGAGTATTCATTTTGAAGCAG 2612
      ::||| |||  ::||| ||| |||
1299 .....AlaThrLeuThrAlaGluSergI 1306
2613 CCGCTTTACCGGCA.....ATCAGCGCGCGCAAGATA 2647
      ::||| |||  ::||| ||| |||
1306 IyIySLeuThrThrGluThrGlySerSerIleThrSerSerAsnGlyGlnT 1323
2648 CGGCAATTAACCTTAAGACAGCAGGATGAGCGCTGCGACGACGGA 2697
      ::||| |||  ::||| ||| |||
1323 hrThrLeuThrAlaIySAspSer.....SergIle 1332
2698 TTAGGCAATTTAAACCTTGACAAAGCCACCATTTACATCAATTCGCGCTA 2747
      ::||| |||  ::||| ||| |||
1333 AlaGlyAsnIleAsnAla.....AlaAsnValIyThrLeuAsnThr..... 1345
2748 TCGCCACAGATGGCGGACGAGGCGCAACCGGACGTGGACAGATGGCGCGC 2797
      ::||| |||  ::||| ||| |||
1346 .....ThrGlyThr..... 1348
2798 GCGCGCGTTCGCGCGCTGCGCGCGTTCCTATTTACGCTACACGCGCA 2847
      ::||| |||  ::||| ||| |||
1349 .....LeuThrThrThrIySAsp 1354
2848 ACTTCGTAGAAATCCGTTTCAACACGCTGACGGTAAACGCGAAATTGA 2897
```



```

1355 SerIlylleasnlathrserglythrleuthrileasnlalys..... 1369
2898 CGGTACGGAGAACATTCCTGTTATGTGGAAGCTTCCTGCGTACCCGACGG 2947
1370 .....Aspa 1371
2948 ACAATTGAAGCTGGCGGAAAGTTCGGAAGCACTTACACCTTGGCGGCG 2997
1371 lalysleuaspglyalaalaserglyasparthrvalalasnalaatr 1387
2998 AACAAATACGGCAACGACCTGCACACCTCGAACAATTCAGCTAGTGA 3047
1388 AsnAlSerGlyser..... 1392
3048 AGGAAAAAGACAAACCGCTGTCGGAACCTTAATTCACCTGCAAA 3097
1393 .GlyAsnValThrAlalysThrserSerSerValasnIlethrGlyaspL 1409
3098 AGCAACACGTGATGCCGGCGGTGACCACTCATCCGCAAGAC 3147
1409 euasnthrIleasnGly.....LeuasnIleIleSerGlyAsn 1421
3148 GGGAGATTCCTCGCATATATCCGTCAAAAGAACAGACTTTCGACAA 3197
1422 Gly.....ArgAsnThrVal.....Ar 1427
3198 ACTGCGGACGAGGACCAAGCCGGAAGGCGGGAAGCAACAGCCGCAAA 3247
1427 glenAargGlyLysIleThrleaspValIlystrIleGlnProGlyAlaLAs 1444
3248 GCGTTCACGGCTGATGCGCGCGCGCGCGATGCGTCAAAAGACAGAA 3297
1444 erylalGluGluValIleGluAlalysArg...ValleuGluLysValLys 1459
3298 AGCGTTGCCGACCGCGCGCGGACGAGCGGGAAGAAATGTCGCAATAT 3347
1460 AspLeuSeraspGluGluArgGluThrLeuAlalys...LeuGlyAlaLs 1475
3348 GCAGCGCGGAGAGAGAAAAAACGCGTGCAGCGGATTAAGACACGCGCT 3397
1475 rAlaVal.....ArgPheValGluProAsnAsnAlaIleThr 1488
3398 TGCGGAACAGCGGAGCGGAAACCGCGCGGCTAC 3435
1488 alAsnThrGlnAsnGluPheThrThrLysProSerSer 1500

seq_name: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:US-08-719-641-10
seq_documentation_block:
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:

```

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

```

```

alignment_scores:
Quality: 295.50 Length: 1196
Ratio: 0.505 Gaps: 62
Percent Similarity: 48.913 Percent Identity: 20.819

```

```

alignment_block:
US-09-303-518D-649 x US-08-719-641-10 ..

```

```

Align seg 1/1 to: US-08-719-641-10 from: 1 to: 1529

```

```

289 TCAGTACGCGCGCTGCGCGGCAATGTTGGCGCATATATATGAGCGT 338
:::|||||::: ||| ::|
507 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 523
::: |||
339 GGCACATACGCGC.....GCGTATACACGTTGATTTGGTGGCG 379
|
523 lAsnLysaspLysGlnPheArgPheAsnAsnValSerIleasnGlyT 540
::: |||
380 AAGGAGAAGATCCCGATCAACATGCTTACTTATTAATTTGGAACGG 429
|||::: |||
540 hrGlyLysGly.....LeuLysPhe.....IleAlaasnGln 550
|||
430 AATAATTATTAAGGAGGAGCTAAAGCCATCCTTATGCGCGCATATCA 479
|||||::: |||
551 AsnAsnPheThr.....HisLysPheaspGlyGluLeuAs 562
|
480 TATGCGCGGTTGCATTAATTTGTACAGATGCAGAACCTTTGAATGA 529
;
562 n..... 562
530 CCAGTTATATGATGGCGGAAATATATCATCAATTAATTAATTAATTAAT 579
::: |||
563 .....IleSerGlyIleValThrIleasnGlnThrThrLysLysasp 576
|||
580 CGTGTTCGATATGGCGGAGGAGCAATATGCGCATCTGATGAAGATGA 629
::: |||
577 .....ValLysTyrrItrPasnAlaSerLysaspSe 586
630 GCCCATATAACCGGAAAGT...TCATATCATATTCGAAGTGGATTTCTT 676
|||
586 rTyrrPasnValSerSerLeuThrLeuasnThrValGlnLysPheThrP 603
GTTGTCGCAAT.....GTTGTCGCAAT..... 693
677 GGCCTC..... 693
603 heLleLysPheValaspSerGlySerAsnSerGlnaspLeuArgSerSer 619
:::
694 .....ACCTTGACAA.....AATGATCATGAGTGTGTCGACA.. 726

```

```
620 ArgArgSerPheAlaGlyValHisPheasnGlyIleGlyLysThrAs 636
727 .GTCACTTAGGTAGTGA.....AAATTAACATATGAC. 759
636 nPheasnIleGlyAlaAsnAlaLysAlaLeuPheLysProAsna 653
760 .....CCATATGGTTTTTACCA..... 777
653 LaAlaThrAspProLysGluLeuProIleThrPheAsnAlaAsnIle 669
778 .....ACAGAGGCTCATTTGGCGACATGGCTCACCATTTATCTA 821
670 ThrAlaThrGlyAsnSer.....AspSerSerValMetPheAspIle 683
822 TGATGCCCAAAAGCAAAAGTGGTTA.....ATTATGGGGTAT 859
683 eHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetLaspSerIleA 700
860 TGCAAACGGGC.....AACCCCTATATAGCAAAAGC 891
700 snIleThrGlyLysAspPheSerIleThrSerHisAsnArgAsnSer 716
892 AATGGCTTCACCTGGTTCGT..... 912
717 AsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThrGlySerAs 733
913 .....AAAGATTGGTTCATGATGAATCTTTGGCTG 943
733 nPheSerLeuLysGluThrLysAspSerPheThrAsnGluIlyrSerLysH 750
944 GAGATACCATTCAGTATTTCTACGACCAAGTCAAAATGGGAATACTCT 993
750 IsAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyLysAsnValThr 766
994 TTACAGCAGATATATATGCG.....ACAGGAAATCAATCAACCCAA 1034
767 LeuGlyGlyLysAsnSerSerSerSerIleThrGlyAsnIleAsnIleTh 783
1035 ACATAGACACAAAT.....TCCTGCCTATTA 1060
783 rAsnLysAlaAsnValThrLeuGluAlaAspThrSerAsnSerAsnThrG 800
1061 GATTAAAAACGACGACCTTCATTTGTTTATGTTCTTTATCCGAGACA 1110
800 LysLeuLysLysArgThrLeuThrLeuGlyAsnIleSerVal..... 813
1111 GCAGAGAACTGTTTATCATGCTGCAGCTGTCACACAGTTATCGACC 1160
814 .....GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGl 828
1161 CAGACTGATATATGAGAAAATATTCTTATTATTCAGCAAGCAAAAGCGC 1210
828 yAsnLeuSerIleAlaGluAspSerThrPheLysGlyLysAlaSerAspa 845
1211 AATTCATCT.....ACCAGCATCATCAAT 1236
845 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 861
1237 CAAGTGTCTGAGATATATTTCAGAGAGATTTTACGGTCTCGCCTGA 1286
862 IleLysGlnGlyValValLysLeuGlnGlyAspIle..... 873
1287 AAATATACCAAACTTGCGACGCGCGGCTTCATATCAGTGAAGACAGTA 1336
874 .AsnAsnLys.....GlyGlyLysAsnIleThrThrAsnAlas 886
1337 CCGTTACTTGGAAA.....GTAAACGGC..GTGGCAACGACGCGCTG 1377
886 ergLysThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGlu..... 900
1378 TCCAAAATCGCGCAAGCGACGCTG.....CAGCTTCAGGCCAAGG 1418
```

```
901 .....LysGlyAspLeuAsnIleLysAsnIleLysAlaAspAl 913
1419 GGAAAAACCA.....GGCTGATCAGCGTGGCGAGGTACAGTCAATT 1462
913 agLuiIleGlnIleGlyLysAsnIleSerGlnLysGluGlyAsnLeuThrI 930
1463 TG..... 1464
930 IeserSerAspLysValAsnIleThrAsnGlnIleThrIleLysAlaGly 946
1465 .....GATCAGCAGGACAGCATGAAGCAAA 1493
947 ValGluGlyLysArgSerAspSerSerGluAlaGluAsnAlaAsnLeuTh 963
1494 ACAAGCCTTAGTGAATCGGCTTG.....GTACGCGCA 1528
963 rIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerLyp 980
1529 GGGGTACGGTCACTGAATCGCATAT..... 1557
980 heAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThrIleGly 996
1558 .....CAGTTCACCCCGACAACCTATTTGCGG...TT 1589
997 AsnAlaSerGlyLysAlaAlaAspAlaLysValThrPheAspLysVa 1013
1590 TCGCGGGGACGTTGGATTTAAACGGCATTCGCTTCGTCACCGTA 1039
1013 LysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsnSerG 1030
1640 TTCAAATACCGATGAAGGGCGATGTTGTCAACACCAATCAAGCAAA 1689
1030 LuValLysThrSerAsnGlySerSer.....AsnAlaGlyAsnAspAsn 1044
1690 GAATCCACCGTTACCATTTACAGCAATTAAGATTTGCTACAAACCGCAA 1739
1045 SerThrGlyLeuThrIleSerAla...LysAspValThrVal..... 1057
1740 TAACACAGCTGGATGATGCAAAAGAAATTCGCTCAACAGCTTGTTG 1789
1058 .AsnAsnAsnValThrSerHisLysThrIleAsnIleSerAlaAlaAlaG 1074
1790 GCGGAAAGATACGACCAACAAACGAGGGCGCTCAACCTGTTTACAG 1839
1074 LysAsnVal...ThrThrLysGluGlyThrThrIleAsn..... 1085
1840 CCGCGCGCAACAGACCGACCTGCTGCTTCGCGGGAACAATTTAA 1889
1086 ...AlaThrThrGlySerValGluValThrAlaGlnAsnGlyThrIleL 1101
1890 CGGCAACATCAGCAACAAACGCAAAAGCTGTTTCAGCGGACGACCA 1939
1101 sgLysAsnIleThrSerGlnAsn.....ValTr 1110
1940 CACCGCAGCGCTCAATCATTTAAAGACCATTTGTCGCAAAAGAGGCG 1989
1110 hrValThrAlaThrGluAsnLeu..... 1117
1990 ATTCTCGCGGGGAATCGTGTGGGACAACGACTGATACCGGCATTT 2039
1118 .....ValThrThrGluAsnAlaValIleAsnAlaThrSe 1129
2040 TAAAGCGAATACTTCCAAATTTAAAGCGACAGCGGGGTTCGCCGA 2089
1129 rGlyThrValAsnIleSerThrLysThrLysAsp.....IleLysGlyG 1144
2090 ATGTTGCCAAATGAAGCGGATTTGCAATTTAGCAATCAGCCCAAGCA 2139
1144 LysIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThr 1160
2140 GTTTTGGTGGCAGCGCATCAAGCCCAACAATCTGATACAGTTCCGA 2189
1161 Leu...LysValSerAsnIleThrGlyLysAspValThrValThrAlas 1176
```

2190 CTGGAGCGCTCTACAAATTGTCTCGAAAAACCATTAACGAGTAAG 2239
 1176 palagialaleuthrthrthralaglyserthrileserthrilethrthg 1193
 2240 TGATTGCTCATG...ACTAAGACC...GACATCAGCGCAATGTGAT 2283
 1193 lysnalalaenliethrthrlysthrlyaspilasnlylgsvalglu 1209
 2284 CTGGCCGAT...CAGGCTCATTTAAATCTCACAGGCGCTGGCCACTCA 2330
 1210 SerSerSerlyserValThrLeuValAlaThrGlyAlaThrLeuAla 1226
 2331 CGGCAATCTAGTCAAAATGGCATACA.....CGTT 2362
 1226 lgllysnliieserlyasnthrValThrIleThrAlaSPserGlyLys 1243
 2363 ATACAGTCAACCAAGCGCCCAAAACGCAACCTTAGCCTGTGGC 2412
 1243 eutThrSerThrValGlySerThrIleasnGlyThrAsnSerValThrThr 1259
 2413 AATGCCACCAACATTTAATCAAGCCACTTAAGCGCAACATCGC 2462
 1260 SerSerInserGlyaspIleGlyThrIleSerGlyasnThr..... 1274
 2463 TTGGGGCAATGCTCATTTAATCAAGCGCAACGCGTCAAAAGGCA 2512
 1275ValasnValThr.....AlaSerThrGly 1283
 2513 GTCTGACGCTTTCCGCGCAACGCTAAGCAAAAGTCCGCACTC 2562
 1283 spleuthrIleGlyAsnSerAlaIlysnValIlysnValIlysnValIly 1298
 2563 AAGGTAATGCTCTCCAGCGAATAGCGATATTCATTTGAAAGCAG 2612
 1299AlaThrLeuthrAlaGlySerG 1306
 2613 CCGCTTACCGGACA.....ATCAGCGCGGCAAGGATA 2647
 1306 ylyleuthrThrGlnThrGlySerSerIleThrSerSerAsnGlyGln 1323
 2648 CGGATTAACCTTAAGACAGCAATGACGCGCTGACGACGAGCA 2697
 1323 hrrthrleuthrAlaIlysnSer.....SerIle 1332
 2698 TTAGCAATTTAAACCTTGACAAAGCCACCATTAACCTTCCGCTA 2747
 1333 AlagIlysnIleasnAla.....AlaasnValThrLeuasnThr..... 1345
 2748 TCGCCACGATGCGGAGGCGCAACCGGACGTGCGACAGATGCGCGC 2797
 1346ThrGlyThr..... 1348
 2798 GCGCGCTGCGCGCGCTTCCGCGCGCTCCATATCCGTTACACCGCA 2847
 1349LeuthrThrThrGlyAsp 1354
 2848 ACTTCGAGTAATCCGTTTCACAGCCTGACGCTAAGCGCAATTTGAA 2897
 1355 SerIlysnIleasnAlaThrSerGlyThrLeuthrIleasnAlaIlysn 1369
 2898 CGGTACGAGCAATTCGCTTATGTGGGAACCTTGCGGCTACGCGCGC 2947
 1370AspA 1371
 2948 ACAATTTGACCTGGCGGAAGTTCCGAAGCATTAACCTTGGCGCTC 2997
 1371 lalysleuaspGlyAlaIlaSerGlyaspThrValValasnAlaThr 1387
 2998 AACATATCCGCGCAACGAACTGCAAGCCTCGAACAAATTTGACGTAGTGA 3047
 1388 AsnAlaSerGlySer..... 1392

3048 AGAAGAACACAAACCGCTGTCCGAAACCTTAATTCACCTCGCAA 3097
 1393 .GlyasnValThrAlaIlysnThrSerSerSerValIasnIlethrGlyasn 1409
 3098 AGCAACACGTGATGCGCGCGCTTGCGCTTACCACTCATTCGCAAGAC 3147
 1409 euAsnThrIleasnGly.....LeuasnIleIleSerGlyasn 1421
 3148 GCGGAGTTCGCCCTGCATATTCGCTCAAGACAAAGAGCTTTCGACAA 3197
 1422 Gly.....ArgAsnThrVal.....Ar 1427
 3198 ACTCGGCAAGCGCAGAAAGCCAAACAGCGCAAAAGACACCGCAAA 3247
 1427 glenArgGlyLysGluIleaspValIlystYrIleGlnProGlyValAlas 1444
 3248 GCCTTGACGCGCTGATTCGCGCGCGCGGATGCCGTGCAAAAGACGAA 3297
 1444 ervaIgluIlyValIleGluIlyasnArg...ValleuGluIlyValIlysn 1459
 3298 AGCGTTGCCGAACGCGCGCGCAGCAGCGGGAATGTGCGCATAT 3347
 1460 AspleuSerAspGluArgIlyThrLeuAlaIlysn..LeuGlyAlas 1475
 3348 GCAGCGGAGGAGAGAGAAAAACGGGTCCAGCGGATTAAGACACCGCT 3397
 1475 rAlaVal.....ArgPheValIgluProAsnAlaIleThrV 1488
 3398 TGGCGAAMACGGCGGAGCGGAACCGCGCGCGTAC 3435
 1488 alAsnThrGlnasnGluPheThrThrIlyProSerSer 1500

seq_name: /cgn2_6/plodata/1/aa/bb_comb.pep:us-08-460-269C-6
 seq_documentation_block:
 : Sequence 6, Application US/08460269C
 : Patent No. 6197548
 : GENERAL INFORMATION:
 : APPLICANT: CLARE, JEFFREY J.
 : ROMANOS, MICHAEL A.
 : TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
 : NUMBER OF SEQUENCES: 17
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
 : STREET: 2200 Clarendon Blvd., Suite 1400
 : CITY: ARLINGTON
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22201
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/460,269C
 : FILING DATE: 02-Jun-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Lebovitz, Richard M.
 : REGISTRATION NUMBER: 37,067
 : REFERENCE/DOCKET NUMBER: Popov-2
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 243-6333
 : TELEFAX: (703) 243-6410
 : INFORMATION FOR SEQ ID NO: 6:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 922 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 : US-08-460-269C-6

alignment_scores:

Quality: 295.00 Length: 1083
Ratio: 0.602 Gaps: 52
Percent Similarity: 45.245 Percent Identity: 22.253

alignment_block:

US-09-303-518D-649 x US-08-460-269C-6 ..

Align seg 1/1 to: US-08-460-269C-6 from: 1 to: 922

```
1417 GGGGAAAACCAAGCGTCGATTCAGCTGGGCGACGCTAGCATCTTTGGA 1466
|||||:|||||:|||||:
45 G|G|U|A|G|G|I|N|I|S|G|Y|I|E|H|I|S|I|E|..... 53
1467 TCAGCAGGACGACGATTAAGGCAAAAACAGCCCTTAGTGAATCGCT 1516
|||||:|||||:|||||:
54 .LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIleL 70
1517 TGTGACGCGCGGCGGTACGTCGCAACGTAATGCCGATATCAGTTCAAC 1566
|||||:|||||:|||||:
70 yValSerGlyArgGlnAlaGlnGlyValLeuLeuGlnAsn..... 83
1567 CCCGACAACTCTATTTGCGCTTTGCGCGCGACGCTTTGATTTAAACG 1616
|||||:|||||:|||||:
84 ...ProAlaAlaGlnLeuArgPheGlnAsnGlySerValThrSerSerG 99
1617 GCATTCGCTTTCGTTCCACCGCATTCAAATACGATGAAGGCGCATGA 1666
|||||:|||||:|||||:
99 Y.....GlnLeuPheAspGlnGlyVal.... 106
1667 TTGTCAACCAATCAAGACAAGAAAGATCCAGCTTACCATTTACAGGCAAT 1716
|||||:|||||:|||||:
107 .....ArgArgPheLeuGlyThrValThrValLysAlaGly 118
1717 AAA.....GATATTGCTTCAACCGGCAATAACACAGCTTGATGATG 1757
|||||:|||||:|||||:
119 LysLeuValAlaAspHisAlaThrLeuAlaAsn..... 129
1758 CAAAAAGAAATTTGCTTCAACGCTTGTGTCGGGAGAAATATGACCA 1807
|||||:|||||:|||||:
130 .....ValSerAspThrArgA 135
1808 AAACGACGCGCGGCTCAACCTTGT.....TACCACCGCGCC 1845
|||||:|||||:|||||:
135 sPaSpAspGlyTlLeuAlaLeuTyrValAlaGlnGlnAlaGlnAlaSer 151
1846 GCAGAAAGCCGACCTGCTGCTTCCGCGGAAACAAATTTA.....AA 1889
|||||:|||||:|||||:
152 TlLeuAlaAspSerThrLeuGlnGlyAlaGlyValArgValGlnArgG 168
1890 CGGCAACATCACGCAAAACAGGCAACTGTTTTCAGCGGC..... 1932
|||||:|||||:|||||:
168 yAlaAsnValThrValGlnArgSerThrTlLeuValAspGlyGlyLeuHis 185
1933 .....AGACCAACACCGCAGCGCTTACATTTAAACGACCAT 1971
|||||:|||||:|||||:
185 lEgLyThrLeuGlnProLeu..... 191
1972 TGTGCGCAAAAAGAGGCGATTCCTCGCGGGAATTCGTGGGCAACGA 2021
|||||:|||||:|||||:
192 .....GlnProGlnAspLeuProSerArgValAlaLeuGlnAspThr 206
2022 CTGATCAACCGC.....ACAT 2038
|||||:|||||:|||||:
206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValAlaPheValP 223
2039 TTTAAAGCGGAAACTTCCAAATTTAAAGCGGACAGCGGTGTTTCCCGC 2088
|||||:|||||:|||||:
223 heGlyAlaAsnGlnLeuThrValAspGlyGlnHisIleThrGlnGlyArg 239
2089 AATGTTCGCAAAAGTGAAGCGGATTTGCGATTTTGACATCAACGCCCAAGC 2138
|||||:|||||:|||||:
240 AlaAlaGly.....ValAlaAl 245
2139 AGTTTGTGTGTCGACCGCATCAAGCCACACAACTGTGACAGTTCCG 2188
|||||:|||||:|||||:
245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262
2189 AC..... 2190
262 sPaLaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyAla 278
2191 .....TGACGCGGTGACAAA 2207
279 ValProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyVal..... 293
2208 TTGTGCGAAAAAACCATTCACGACGATAAGATGATGCTTCATGTGACTA 2257
|||||:|||||:|||||:
293 ..... 293
2258 AGACCGCATCAGC...GCCAATGTCATCTTGCCGATCAGCTCATTTA 2304
|||||:|||||:|||||:
294 .....AspValSerAspSerThrValAspLeuAlaGln..... 304
2305 AATCTCACAGGCGCTTGCCACACTCAACGCAATCTTAGTGCATAATGGCGA 2354
|||||:|||||:|||||:
305 SerIleValGlnAlaProGlnLeuGlyAlaAlaIleArgAlaGlyArgG 321
2355 TACACGTTATACAGTACGACCAAC.....GCCACCAAAACGGCAAC 2398
|||||:|||||:|||||:
321 yAlaArgValThrValSerGlySerLeuSerAlaProHisGlyAsnV 338
2399 TTAGCCCTCGTGGGCAATGCCCAAGCAACATTTAAT..... 2433
|||||:|||||:|||||:
338 allEgLyThrGlyGlyGlyAlaArgArgPheProProAlaSerPro 354
2434 ...CAAGCCACATTTAAACGCGCACATCGCTTCCGCAATGCT...TC 2477
|||||:|||||:|||||:
355 LeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgValLeuLe 371
2478 ATTTAATCTTAAGCGACCGCGCTACAAAAGCGAGCTTGACGCTTCCG 2527
|||||:|||||:|||||:
371 uTyrArgValLeuProGlnProValLys.....LeuThrLeuAlaG 385
2528 GCAAGCTTAAAGGCAAGGTAAAGCATTCGACATCAACGTTATGTCTCC 2577
|||||:|||||:|||||:
385 lGlyAlaGlnGly.....GlnGlySerPleVal 394
2578 CTAGCGATTAAGCAGATTCATTTTGAAGCAGCGGCTTTACCGGACA 2627
|||||:|||||:|||||:
395 AlaThr.....GluLeuProProIleProGlyAl 404
2628 AATCAAGCGGCGGCAAGATTCAGGCTTACACTTAAAGACAGGAATGA 2677
|||||:|||||:|||||:
404 aSerSerGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTyrP 420
2678 CGCTGCGCTGACGACGAGATTTAGCAATTTAAACCTTGAACAGCGCAC 2727
|||||:|||||:|||||:
420 hr...GlyAlaThrArgAlaValAspSerLeuSerIleAspHisAlaThr 435
2728 ATTACACTCAATTCGCGCTATTCGCCACGATGCGGCGGCGCAACCGG 2777
|||||:|||||:|||||:
436 TrpValMet..... 438
2778 CAGTGGCAGAGATGCGCGCGCGCGCTTCGGCGCTTGCCCGCTTCC 2827
|||||:|||||:|||||:
439 .....ThrAsp.....AsnSerAsnValGlyA 446
2828 TATTATCGCTTACACCGCCCACTTCGTTAGAA..... 2859
|||||:|||||:|||||:
446 lLeuArgLeuAlaSerAspGlySerValAspPheGlnInProAlaGln 462
2860 ...TCCCGTTTCAACAGCTGACGGTAAAGCGCAATTTGACGGTCCAGCG 2906
|||||:|||||:|||||:
```

```

463 AlaGlyArgPheLysValLeuMetValAsp...ThrLeuAlaGlySerGI 478
2907 AACATTCGGCTTTATGTGGAAGCTCTTCCGGCTACCCGACGACAAATTGA 2956
478 yLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuV 495
2957 AGCTGGGGGAAAGTTCGGAAGGACCTTACACCTTGGCGGTGCAACATATAC 3006
495 AlValMetAlaGAspAlaSerGlyGlnHisArgLeuTrpValAlaGAsnSer 511
3007 GGCAGACGACCTGCAAGCCTCGAACAATTGACGCTAGTGGAGAGAAAGA 3056
512 GlySerGlnProAlaSerGlyAsnThrMetLeuValGln..... 525
3057 CAACAAACCGCTGCGAAACCTTAATTCACCTGCGCAAAAGCA..... 3102
526 ....ThrProArgGlySerAlaAlaThrPheTrpLeuAlaAsnLysAspG 541
3103 ..CACGTGATGCGCGGCGCTGGCGTTACCACTCATCCGCAAAAGAGCGC 3150
541 LysValAspIleGlyThrTyrArgTyrArgLeu..... 552
3151 GAGTCCGCTGCATTAATCCGGTCAAGAACAGAGCTTCCGACAAACT 3200
552 ..... 552
3201 CGGCAAGCAGAGCCAAACAAACAGCGGCAAAAGACAGCGCAAGGC 3250
553 .....AlaAlaAsnGlyAsnGlyGlnTrp... 560
3251 TTGACGGCTGATTCGGCGCGCGCGATGCGCTGCAAAAGACAGAAAG 3300
561 ....SerLeuValGlyAlaLysAlaProProAla.....ProLysPr 573
3301 GTTCCGCAACCGCGCGCGAGCAGCGGCGGAAATGTGGCATTTATGCA 3350
573 AlaPro...GlnProGlyProGlnProGlyProGln..... 584
3351 GCGCGAGGAAGAGAAACGCGGTGCGAGCGGATTAAGACCGCGCTTG 3400
585 .....ProPro... 586
3401 CGAAGACGCGGAAGCGAAACCGCGCGGTACACCGGCTTCCCGCG 3450
587 .....GlnProProGlnProProG 593
3451 GCCCGCGCGCGCGCGCGGATTCGCCCACTGCAACCCCA..... 3492
593 InProProGlnProProGlnProProGlnProGlnProAlaProAl 609
3493 ..CGCAGCGCCCAACCGCGAGCGAGCTGATCAGCGCTTATGCCAATAGC 3541
609 aProGlnProProAlaGlyArgGlyLeu...SerAlaAlaAlaAsnAla 625
3542 GTTTGAGTGAATTTCCGCCACGCTCAACAGCGT.....TTCGCCGTA 3585
625 laValAsnThrGlyAlaGlyLeuAlaSerThrLeuTrpTyrAlaGln 641
3586 CAGACACGAATTAGACCGCGTATTCGCCGAAGACCGCGCGCAAC..... 3627
642 SerAsnAlaLeuSerLysArgLeuGlyGlnLeuArgLeuAsnProAspAl 658
3628 ....GCCGTTTGACAGCGCATCGGAGACCAACAACACTACCGCTGCG 3673
658 aGlyGlyAlaTrpIleArgGlyPheAlaGlnArgGlnGlnLeuAspAsn 675
3674 AAGATTTCCGCGCTACCGCAACAAACGACGCTGCGCAATCGTATG 3723
675 rGAlaGlyArgArgPheAspGlnLysValAlaGlyPheGlnLeuGlyAla 691
3724 CAGAAAAACCTC.....GCCAGCGCGCGCGCTGCGCATCCGCTGT 3761
692 AspHisAlaValAlaValAlaGlyLysArgTrpHisLeuGlyLysVal 708

```

```

3762 TTCGACAAACCGACGAAACACCTTC...GACGAGCATCGGCAACT 3808
708 aGlyTyrThrArgGlyAspArgGlyPheThrGlyAspGlyGlyLysIst 725
3809 CGGACGCGCTTGCCCGCGCGCGCTTTCGGGCAATACGGCATCGACAG 3858
725 hrAspSerValHisValGlyLysTyrAlaThrTyrIleAlaAsnSerGly 741
3859 TTCATCATCGGCATCAGCGCGCGGCTTTTACGACGGGACAGCT... 3906
742 PheTyrLeuAspAlaThrLeuArgAlaSerArgLeuGlnAsnAspPheLy 758
3907 .....TCAAGACGC.....ATCGAGGCAAAATCCGCGCGCG 3940
758 sValAlaGlySerAspGlyTyrAlaValLysGlyLysTyrArgThrHisG 775
3941 TGCTGCATTCAGGCATTCAGCAGATACCG...GCCGTTTCGGCGGA 3987
775 LysValGlyValSerLeuGlnAlaGlyArgArgPheAlaHisAlaAspLy 791
3988 TTCGGCATCGAAGCGCATCGCGCAAGCGCTATTTCGTCGAAAGAC 4037
792 TrpPheLeuGlnProGlnAlaGlnLeuAlaValPheArgValGlyGly 808
4038 GATTAACCGCTACGAAAC...GTCAATATCGCAACCCCGCGCTTCAT 4084
808 yAlaTyrArgAlaAlaAsnGlyLeuArgValArgAspGlnLysIst 825
4085 TCACCGCTACCGCGCGCGCATTAAGCAGATTTATTCATCAAAACCGCG 4134
825 ervaLeuGlyArgLeuGlyLeuGlyValGlyLysArgGlyLeuAla 841
4135 CAACATTTCCATCAGCGCTTATTGAGCTGTCTTATACGATGCGCG 4184
842 GlyGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnLpH 858
4185 TTCGGGCAAAAGTCGACACGCGCTAATACCGCGTATTCGCTACGATT 4234
858 eaSpGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGlu 875
4235 TCGCAAAACCGCGAGTCGGAATGGCGGCTAAGCGCAATCAAGAGT 4284
875 euArgGlyThrArg...AlaGlyLeuGly..... 883
4285 TTCAGCGTCTCCCTCCAGCTGCGCGCGCAAGGCGCAACTGGAAGC 4334
884 ....LeuGlyMetAlaAlaLeuGlyArgGlyHisSerLeuTyrAl 898
4335 G.....CAACAGCGCGCGCATTAATTAAGCTACCGCTGG 4371
898 aSerTyrGlyTyrSerLysGlyProLysLeuAlaMetProTrp 912

seq_name: /cgn2_6/plodata/1/laa/5b_comb.pep:us-08-617-697-10
seq_documentation_block:
: Sequence 10, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matlare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

675 heilLeuysheValaspserrGlySerAsnGlyGlnaspreuayrSer 691
694ACSTTTGCACA.....AATGATACAGTGGTGGCACA..726
692 ArgArGserPheulagIyValHisPheAsnGlyIleGlyGlyAsThr 708
727 ..GTCAACTAGGTAGGA.....AAATTAACATAC..759
708 nPheAsnIleGlyAlaAsnAlaLeuAlaLeuPheUlysPheProAsnA 725
760CSATATGGTTTTTACA.....777
725 laAlaThrAsprouLysGluLeuProIleThrPheAsnAlaSnIle 741
778ACGAGAGGCTATTTGGCCAGACAGTGGCTACCAATGTTATGTA 821
742 ThrAlaThrGlyAsnSer.....AspSerValMetPheAspI1 755
822 TGATGGCCCAAGCAAAAGTGGTA.....ATTATGGGGAT 859
755 enHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSerIleA 772
860 TGCAAACGGC.....AACCCCTATATAGAAAAGC 891
772 snIleThrGlyLeuAspPheSerIleThrSerHisAsnArgAsnSer 788
892 AATGCTTCCACTGGTGTGT.....912
789 AsnAlaPheGluIleUlysAspreuThrIleAsnAlaThrGlySerAs 805
913AAGATGGTTCATGATGAATCTTGCTG 943
805 nPheSerIleUysGlnThrLysAspSerPheGlyAsnGlyUysThrL 822
944 GAGATACCATTCAGTATTTAGACCAACAGTCAAATATGGGAATACCT 993
822 laAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThr 838
994 TTTAACGACGATATATATGCG.....ACAGAAAAATCATATGCCAA 1033
839 LeuGlyGlyGluAsnSerSerSerSerIleThrGlyAsnIleAsnIleTh 855
1035 ACATGACACCAT.....TCTCGCTATTA 1060
855 rAsnLysAlaSnValThrLeuGlnAlaAspThrSerAsnSerAsnThrG 872
1061 GATTAAAAACAGCAACCGTTCAATTGTTAATGTTTATCCGACAGACA 1110
872 LyLeuLysArgThrLeuThrLeuGlyAsnIleSerVal.....885
1111 GCAAGAGAACCTGTTTATCATCTGCACAGTCGTCGAACAGTTATGACG 1160
886GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValG1 900
1161 CAGCTGATATATGAGAAAATATTTCTTATTTAGACAGGAAGAGCG 1210
900 yAsnLeuSerIleAlaGluAspSerThrPheUysGlyGlnAlaSerAspA 917
1211 AATTACTACTTACAGCAACATC...AATCAAGAGCTCGAGAGATTATAT 1257
917 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 933
1258 TTCAGAGAGATTTACGCTCTCGCCTGAATAAATGCAAACTTGGCAGG 1307
934 IleUysGlyValValLysLeuGlyAspIleAsnAsnLys.....946
1308 CGCGGGCGTTCAATACAGTGAAGACAGTACGTTACTTGGAAA.....G 1351
947 ..GlyLeuLeuSnIleThrThrAsnAlaSerGlyThrGlnLysThrIleI 963
1352 TAAACGGC...GTGGCAACAGCACCGCTGTCCAAATATGGCAAGAGCG 1398
963 leAsnGlyAsnIleThrAsnGlu.....LysGlyAsp 973

1463 |||::: :::
3069 GTCCGAAAACCTTAATTCACCCCTGCAAAAGACACACGTCGATGCCGGG 3118
1469 rSerSerValaSnIleThrGlyAspLeuAsnThrIleAsnGly.... 1484
3119 CGTGGCGTTACCACTCATCCGAAAGACGGGAGCTCCGCTGCATAT 3168
1485 LeuAsnIleIleSerGluAsnGly..... ArgAsn 1494
3169 CCGTCAAGAACAGAGCTTCCCAACAACCTCCGCAAGCCGAACCA 3218
1495 ThrVal..... ArgLeuArgGlyIleGluIleAs 1504
3219 AAAACAGCGGAAAAAGACACGCGCAACCTTGACGCGGCTGATTGGG 3268
1504 pValIleThrIleGlnProGlyValAlaSerValGluIleGluVal 1521
3269 CCGGCGCGGATCCGTCGAAAAGACGAAAGGTTGCCGAACCGGCCGG 3318
1521 lAlaArg... ValLeuGluValValIleAspLeuSerAspGluLys 1536
3319 CAGGAGCGGCGGAAATGTCGGCATTTGACGCGGAGAGAGAGAAAA 3368
1537 GluThrLeuAlaLys... LeuGlyValSerAlaVal..... Ar 1548
3369 ACGGCTGACGCGGATAAAAGACACCGCTTGCGCAACGCGGAGCGG 3418
1548 gPheValGluProAsnAsnAlaIleThrValAlaSnThrGlnAsnGluPheT 1565
3419 AAACCGGCGCGGCTACG 3435
1565 hrThrIleProSerSer 1570

seq_name: /cgn2_6/prodata/1/iaa/bh_COMB.pep:US-08-460-269C-2
seq_documentation_block:
Sequence 2, Application US/08460269C
Patent No. 6197548
GENERAL INFORMATION:
APPLICANT: CLARE, JEFFREY J.
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
STREET: 2200 Clarendon Blvd., Suite 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS.
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,269C
FILING DATE: 02-Jun-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lebovitz, Richard M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: Popov-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 243-6333
TELEFAX: (703) 243-6410
INFORMATION FOR SEQ ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 910 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2
alignment_scores:
Quality: 283.00 Length: 1100
Ratio: 0.581 Gaps: 52
Percent Similarity: 44.273 Percent Identity: 21.909
alignment_block:
US-09-303-518D-649 x US-08-460-269C-2 ..
Align seg 1/1 to: US-08-460-269C-2 from: 1 to: 910
1366 AACGACCCCTGTGCAAAATGCGCAAGGACGCTGCAGCTTAAGCCAA 1415
1416 AGGGAAACCAAGGCTCGATCAGCGGCGGAGCGGAGCTGCTATTG 1465
54 nglySerAspProGlyValAlaArgThrAlaSerGlyThrThrIle... 69
1466 ATCAGCAGCGAGAGATTAAGCAAAACAGCTTTAGTGAATCGGC 1515
69 69
1516 TTGCTACGCGGAGGCTACGCTGCACTGATCCGATATCATGTTCAA 1565
70 LysValSerGlyArgGlnAlaGlnGlyIleLeuLeuGluAsn..... 83
1566 CCGGCAAACTGATTTGCGCTTGCGGCGGAGCGGCTTGATTAAGC 1615
84 ProAlaIleGluLeuGlnPheArgAsnGlySerValThrSerSerG 99
1616 GGCATTCGCTTTCGTTCCACCGATTTCAAAATACCGATGAAGCGCGAG 1665
99 Ly..... GluLeuSerAspGlyIle... 106
1666 ATTGTACACCAATCAAGCAAGCAAGATCCCGCTTACCATTTACAGGCA 1715
107 ArgArgPheLeuGlyThrValThrValLysAlaG 118
1716 TAA..... GATATTCTTACACCGGCAATTAACACAGCTTGANA 1756
118 LysLeuValAlaAspHisAlaThrLeuAlaAsnValGly..... 131
1757 GCAAAAAGAAATTCCTTACACGCTTGCTTGCGGAGAAAGATACGACC 1806
132 AspThrTrp 134
1807 AAAACGAGCGGCGCTCAACTTGT..... TACCAGCCCGC 1844
135 AspAspAspGlyIleAlaLeuTyValAlaGlyGlnAlaIleAlaIle 151
1845 CCGAGAGACCGGACCCGCTGCTTCGCGGAGCAAAATTA.....A 1888
151 rIleAlaAspSerThrLeuGlnGlyAlaGlyValGlnIleGluArg 168
1889 ACGGCAACATCAGCAAAACAGGCAAACTGTTTTCAGCGGAGACCA 1938
168 LysAlaAsnValThrValGlnArgSerAlaIleValAspGlyLeu... 183
1939 ACACCCGAGCGCTTACATATTAAAGACCAATTGTGCG...CAAAAGA 1985
184 HistIleGlyAlaLeuGlnSerLeuGlnProI 194
1986 GGGCATTCCTCGCGGGAATCGTGTGGAGCAACGACTGATCAACCGCA 2035
194 uAspLeuProProSerArgValValLeu.....ArgA 205
2036 CATTTAAGCGGAAACCTTCAATTAAGGGGAGCGGCTGTTCC 2085

```
205 sphrAsnValThrAlaValProAlaSerGlyAlaProAlaValSer 221
208 CGCAATGTCCTCAAGTGAAGCGCATTTGACATTTGAC.....AA 2126
222 .....ValLeuGlyAlaSerGluLeuThrLeuAspLysI 233
2127 TCAC.....GCCCAAGCATTTTGGTGGTGG 2152
223 yHisIleThrGlyLysArgAlaIleGlyAlaAlaIleMetGlnGlyAlaV 250
2153 CACCGCATCAAGCCACACCAATCTGTACAGTTGGAC..... 2190
250 aValHisLeuGlnArgAlaThrIleArgArgGlyAspAlaLeuAlaGly 266
2190 ..... 2190
267 GlyAlaValProGlyGlyAlaValProGlyGlyAlaValProGlyLys 283
2191 .....TGCACGGCTGTGACAA 2206
283 eGlyProGlyGlyPheGlyProValLeuAspLysTyrPtyGlyVal.... 298
2207 ATTGTGTGAAAAAACCATTACCAGCATTAAGTGTGCTTCTATTGACT 2256
298 ..... 298
2257 AAGACGACATCAACGGC...AATGTCATCTTGCAGTACGCTCATTT 2303
299 .....AspValSerGlySerSerValGluLeuAlaGln..... 309
2304 AATATCTCAAGGGCTTGCACACTCAACGGCAATCTTAGTCAAAATGGCG 2353
310 .....SerIleValGluAlaProGluLeuGlyAlaIleArgValGlyArg 326
2354 ATACAGCTTTATACAGTCAGCCACAAAC.....GCCACCAAAAGCGCAAC 2397
326 LysAlaArgValThrValProGlyGlySerLeuSerAlaProHisGlyLys 342
2398 CTTAGCTCTGTGGCAATGCCCAAGCAACATTTAATCAAGCC..... 2439
343 ValIleGluThrGlyLysAlaArgArgPheAlaProGluAlaIleProle 359
2440 .....ACATTAAGCGCAACATCGCTGTGGCAATGCT...TCAT 2479
359 uSerIleThrLeuGlnAlaGlyAlaHisAlaGlnGlyLysAlaLeuLeu 376
2480 TTAATCTAAGCACCGCCGCTACAAAACGGCACTGTAGCTTTCCGGC 2529
376 yArgValLeuProGluProValLys.....LeuThrLeuThrGly 389
2530 AAGCTTAAGCAAAAGTAAGCCATTCGCACCTCAACGGTAATGTCTCCCT 2579
390 GlyAlaAspAla.....GlnGlyAspIleValAl 399
2580 AGCCGATAAGCACTATTTCATTTTGAAGAGCGGCTTACCGCAAAA 2629
399 aThr.....GluLeuProSerIleProGlyThrS 409
2630 TCAGCGCGGCAAGATACGATTTACCTTAATAAGACAGCAATGAGC 2679
409 erIleGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTyrThr 424
2680 CTGCGCGTACGACGGAATTAAGCAATTTAAACCTTGCAACGGCACCAT 2729
425 ...GlyAlaThrArgAlaValAspSerLeuSerIleAspAlaThr 440
2730 TACACTCAATTCGGCTATTCGCCACGATGGCGAGGGCGCAACCGGCA 2779
440 pValMet..... 442
2780 GTGCGACGATGCGCGCGCGCGCTTCGCGCTTCGCGCGCTTCCTTA 2829
443 .....ThrAsp.....AsnSerAsnValGlyAla 450

2830 TTATCCGTTACACCGCCCAACTTCGTAGAA..... 2859
451 LeuArgLeuAlaSerAspLysSerValAspPheGlnGlnProAlaGluAl 467
2860 ..TCCCGTTTCAACGCGTACGAGCTAAACGGCAATTCAGCGTCAGGCA 2908
467 aGlyArgPheLysValLeuThrValAsn...ThrLeuAlaGlySerGlyL 483
2909 CATTCGCTTTATGCGGAACCTTCGCGCTACCGCAGGACCAATTTAG 2958
483 eupHeArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuVal 499
2959 CTGCGGAAAGTTCGGAAGCATTACCTTGGCGGTCAACAATATACCG 3008
500 ValMetGlnAspAlaSerGlyGlnHisArgLeuThrValArgAsnSerG 516
3009 CACGACACCTGCAAGCCTCGAACAATTTACCGGTAGTGGAGAAAGACA 3058
516 ySerGluProAlaSerAlaAsnThrLeuLeuValGln..... 529
3059 ACAACCGCTGTGCGAAAAACCTTAATTCACCTTCGCAAAACGAA..... 3102
530 ..ThrProLeuGlySerAlaAlaThrPheThrLeuAlaAsnLysAspGly 545
3103 CACGTGATGCGCGCGCGGTGCGGTATCCACTCATTCGCAAAAGCGCGA 3152
546 LysValAspIleGlyThrTyrArgTyrArgLeu..... 556
3153 GTTCGCGCTGCATATCCGCTCAAAAGAACAGAGCTTCGCAAAACTCG 3202
556 ..... 556
3203 GCAGGCAAGACCCAAACAGCGGAAAGACACAGCGCAAGCTT 3252
557 .....AlaAlaAsnGlyAsnGlyGlnTyr... 564
3253 GACGCGCTGATTCGCGCGCGCGGTATCCGCTCGAAAGACAGAAAGCT 3302
565 ...SerLeuValGlyAla.....Lys 570
3303 TCCGACACGCGCGCGCGGCAAGCGCGGAAATGTCGATTAATGACAG 3352
570 sAlaProProAla..... 574
3353 CGAGGAGAGAAAAACGGGTGACGCGGATTAAGACACCGCTTGGCG 3402
574 ..... 574
3403 AAACAGCGGAAAGCGAAACCGCGCGGCTACACCGCTTCCCGCGC 3452
575 .....ProLysProAla..ProGluProGly..ProGln 584
3453 CCGCGCGCGCGCGCGGATTTGCCCACTGCACACCGCAACCGCAGCGC 3502
585 ProProGluProProGluProGluProGluAla..ProAlaProGluPro 601
3503 AACCGCAGCGGCACTGATCAGCGCTTATGCCAATACGGGTGTGAGTGA 3552
601 roAlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaValAsnThr 616
3553 TTTTCGCGCAGCTCAACAGCGT...TTCGCGTACAGAGCAATT 3596
617 GlyGlyValGlyLeuAlaSerThrLeuTyrIleAlaGluSerAsnAlaLe 633
3597 AGACGCGATTTTGCAGAGACCGCGCAAC.....GCCGTTT 3634
633 uSerLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyAlaT 650
3635 GCACAAAGCGGCACTCGGGAACACCAAACTACCTTCCGCAACATTTCCGC 3684
650 rpeGlyArgGlyPheAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArg 666
```



```

357 ..... 357
171 ILeVallyThrLysGluTrpLeuAspProGluAsnValThrIleG1 187
358 ..... AACACGTTGATTTGGGCGGAGGAAGAATCCG 394
187 uAlaProSerAlaSerArgValGluLeuGlyAlaAsp...ArgAsnSerH 203
395 ATCAACATCGTTTACTATATATAATGGAACGAAATATTAATAAACA 444
203 IsSerAlaGluValIleLysValThrLeuLysLysAsnAsnThrSerLeu 219
445 GGGACTAAAGCCATCCTTATGGCGGATATCATATGCCGGCTTGCA 494
220 ThrThrLeuThrAsn.....ThrThrIleSe 228
495 TAAATTTGTCACAGTCGAGAACTGTTGAATGACAGT..... 534
228 rAsnLeuLysSerAlaHisValIleAsnIleThrAlaArgArgLysL 245
535 ..TATATGATGGCGGAAATATATCGATCAAAATATTAATACCTGACCGT 582
245 eutThrValAsnSerSerIleSerIleGluArgLysSerHisLeuIleLeu 261
583 GTTCGTATGGGCGAGCGCAATATGGCATCTGATGATGAAGATGAGCC 632
262 HisSerGluGlyGlnGlyGlnGlyValGlnIleAspLysAsp.... 276
633 CAATACCCGCAAGTTCATATCATATTTGCAAGTCGATTTCTGGCTCG 682
277 .....IleThrSerGlu..... 280
683 TTGTCGCAATACCTTTGCACAAATGATGAGTGT..... 720
281 ..GlyGlyAsnLeuThrIleTyr.....SerGlyGlyTrpValAspVal 294
721 ..GGCAGCTCAACTAGTAGTGAATAAATAACATAGCCATATGG 767
295 HisLysAsnIleThrLeuGlySer.....G1 303
768 TTTTACCACAGAGAGCGTCATTTGGGAC.....A 799
303 yPheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLys 320
800 GTGCTCAACAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTTAAT 849
320 erGlyArgAsnAsnLeuThrIleThrAlaGln..... 330
850 AATGGGATATTGCAACGGCAACCCCTATATAGAAAAACAATGCTT 899
331 ..GlyThrIleThrSerGlyAsn.....SerAsnGlyPh 341
900 CCAGCTGTCGTAAAGATGTTCTATGATGAATCTTTCCTGAGACATA 949
341 eArg.....PheAsnAsnValSerLeuAsnSerL 351
950 CCATTCAGTATTTACAGACCCAGTCAAAATGGAAATACTCTTTAAC 999
351 eu.....GlyGlyLysLeuSerPheThr 358
1000 GACGATAT..... 1008
359 AspSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 375
1009 .....AATGGCAGAGAAATATCAATCCAAAC 1036
375 spheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysA 392
1037 ATGAACACAATTCCTCGCTAATAGATTAATAACAGAACCGTT..... 1080
392 laProLysValSerTrpPheTyrArgAspLysGlyArgThrLysTrpAsn 408
1081 .....CAATGTTTAAATGTTCTTATC 1103

```

```

409 ValThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAs 425
1104 CGAGACAGCAGAGAAACCTGTTATCATGCTGCAGGTGCTGACACAGT 1153
425 pSerThr.....GlySerGlySerT 432
1154 ATGACCCGAGCTGAATTAATGAGAA.....AATATTTCTTT..... 1191
432 hrcLysProSerIleArgAsnAlaGluLeuAsnGlyIleThrPheAsnLys 448
1192 .....ATGACGAGAGAAAGCGGAATGATTAATTAATTAACAGCA 1229
449 AlaThrPheAsnIleAlaGlnGlySer.....ThrAlaAs 460
1230 CATCAATCAAGTCGCGAGAGATTAATTTCCAGAGATTTTACGCTCT 1279
460 nPheSerIleLysAlaSerIleMetProPheLysSerAsnAlaAsnTyrA 477
1280 CGCCTGMAAATAAGCAAACTGGCAAGCGCGGCTTCATATCAATGATA 1329
477 laLeuPheAsnGluAsp.....IleSerValSerGly 487
1330 GACAGTACCGTACTTGGAAAGTAAACGCG..... 1359
488 GlyGlySerValAsnPheLysLeuAsnAlaSerSerAsnIleGlnTh 504
1360 .....GTCGCAACGACCGCTGTCGCAAAATCGGCAAGGCA 1396
504 rProGlyValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerT 521
1397 CGCTGCAGCTTCAAGCAAAAGGGAACCAAGCTCGATCAGCGTG... 1443
521 hLeuAsnLeuLysAlaGlnGlySerThrGluThrAlaPheSerIleGlu 537
1444 .....GGCGAGCTGACGTCATTTTGA 1466
538 AsnAspLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGluVa 554
1467 TCAGCAGCAGCAGATAA..... 1485
554 lGluGlyThrAspSerArgValAsnLysGlyValAlaIleLysLysAsnI 571
1486 .....GCCAAAAACACCTTTAGT 1506
571 leThrPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrTrp 587
1507 GAATTCGCTGTCACGCGGAGGTACGTCGAATGCAATGCGGATTA 1556
588 GluIle.....LysGlyAsnValThrIleAsnLysAsnTh 599
1557 T.....CAGTTCAACCCGCAACAACTCTATTTCG 1585
599 rAsnAlaThrLeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeuA 616
1586 GCTTTCGCGCGGAGCTTTGATTAACGGCATTCGTTCTGTTCCAC 1635
616 snIleAlaGlyAsnValIleAsn...AsnGlyAsnLeuThrThrAlaGly 631
1636 CGTATTCAAATAACGATGACGAGGCGATGATTGTCACCAACATCAGA 1685
632 SerIleIleAsnIle...AlaGlyAsnLeuThrValSerLysGlyAlaAs 647
1686 CAAGAATCCACCGTTTACCATTAAGCAATTAAGATATGCTACACCG 1735
647 nLeuGlnAlaIleThrAsnTyrThrPheAsn.....ValAlaGlySerP 662
1736 GCATTAACAACAGCTTGGATACCAAAAAAGAAATGCTTACACGCTGG 1785
662 heAspAsnAsnGlyAla...SerAsnIleSerIleAla...ArgGlyGly 676
1786 TTGGGCGAGAAAGATACGACAAACGAGCGGCGCTCAACCTGTTTA 1835

```

677 AlaLysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThr 692
1836 CCAGCCCGCCGAGACGACCGACCTGCTTCCGGCGGACAAAT 1885
692 rAsnSerAspThrThrTyrArgThrIle.....I 702
1886 TAAACGGCAACATCAGCAACAAACGCAACATGTTTTCAGCGGCA 1935
702 IeLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleLysPlys 718
1936 CCAACACCCGACGCTACATATTAAACGACATGGTGGCAAAAGA 1985
719 LysSer...AspIleGluIleGlnIleGlyGlyAsnIleSerGlnLys 734
1986 GGGCATTCCTCGGGGAAATCGTGGGCAACGACTGGATCAACCGCA 2035
734 u.....GlyAsnLeuThrIleSerSerAspLysValAsnIleT 747
2036 CA.....TTTAAACGGCAAAAC 2052
747 hrAsnGlnIleThrIleLysAlaGlyValGluGlyLysArgSerAsp 763
2053 TTCCAAATTAAAGCGGACGCGGCTGGTTCCCAATGTTGCCAAAGT 2102
764 SerGluAlaGluAsnAlaAsnLeuThrIleGlnThrLysGluLeuLysLe 780
2103 GAAAGCGCATTTGGCATTTGAGCAATCAGCCCAACAGATTTTGGTGGC 2152
780 uIaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaL 797
2153 CACCCGCAAAAGCCACACATCTGTACAGCTGTGAGCGGCTG 2202
797 yAsnGlySerAspLeuThrIleGlyAsnAlaSerGly..... 809
2203 ACAATTGTGTGAAAAAACCATTCACGACATTAAGATGCTTCATT 2252
810 GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLys 826
2253 GACTAAGACGACATCAGCGGCAATGTCGATTCGATCAGCTCATT 2302
826 sIleSerThrAsp...GlyHisAsnValThrLeuAsnSerGlu..... 839
2303 TAAATCTACAGGCTTGCACATCAACGCGCAATCTTAGTCAAAATGGC 2352
840ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 851
2353 GAT.....ACACGTTATACAGTACG...CACACGCGACCCCAACGG 2393
852 AspAsnSerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAs 868
2394 CAACCTTAGCCTCGTGGCAATGCCCAACCAATTTAATCAAGCCACAT 2443
868 nAsnValThr.....SerHisLysThrIleAsnIleSerAla 881
2444 TAAACGGCAACATCGGCTTGGGCAATGCTTCAATTAATCTAAGCGAC 2493
881 IaAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThr 897
2494 CACGCGGCTA.....CAAAACGGCACTGACGCTTCCGGGAA 2531
898 GlySerValGluValThrAlaGlnAsnGly.....ThrIleLysGlyAs 912
2532 C...GCTAAGCAAAAGTAAAGCATTCGACATCAACGGAATATCTCC 2578
912 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 928
2579 TAGCGGATTAAGGAGTATTCATTGTAAC.....AGC 2613
928 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 944
2614 CGCTTACCGGAAATCAGCGGCGCAAGATACGGA.....TT 2654
945 ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAsnVa 961

2655 ACATTAAAGACAGCATGAGCGTCCGCTCAGCGACGAATTAGCA 2704
961 IAsnIleThrAlaSerGlyAsnThrIleu.....LysValSerA 974
2705 ATTTAAACCTTGACACGCGACCATTCACATTCCTGCTATTCGGCAC 2754
974 snIleThrGlyGlnAspValThrValThrAlaAsp..... 985
2755 GATCGGCGAGGCGCAACCGGACGTGCGACAGATCCGCGCGCGC 2804
986AlaGlyAlaLeuThrThrAla..... 993
2805 TTGCGCGCGTTCGCGCGCTTCCATATTCGCTTACACCGCAACTCG 2854
994GlySerThrIleSerAlaThrThrGlyAsnAla 1004
2855 TAGAATCCGTTTCACACGCTGACGCTAAAGCAATGAGCGTACG 2904
1005AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1016
2905 GGAACATTCGCTTTATGTGGAACCTTCGCTACCGCAGCAAAAT 2954
1016 1016
2955 GAAGCTGGCGGAAGTTCGAGAGCACTTACCTTGGCGGTACAAAT 3003
1017ValGluSerSerSerGlySerValThrLeuValAlaThrGly 1031
3004ACCGGCAAGCACTGCAAGCTCGAACAATGACGGTA 3042
1031 IaThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAla 1047
3043 GTGAGAGGAAGACAAACACCGCTGTCGAAAACCTTAAT..... 3084
1048 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1064
3085TTCACGCTGCMAA 3097
1064 nSerValThrThrSerSerIleSerGlyAspIleGluGlyThrIleSerG 1081
3098 ACGAACACGTCGATCGCGCGCGGTTCACACTCATCCCAAGAC 3147
1081 LysAsnThrValAsnValThrAla.....SerThr 1090
3148 GCGGAGTTCGCGCTGATTAATCCGCTCAAGAACAGCTTCCGACAA 3197
1091 GlyAspLeuThrIleGlyAsnSerAlaLys.....ValGluAlaLys 1104
3198 ACTCGGCAAGGCAAGCCAAACCAAGCGCGGAAAGAACAGCGCA 3247
1104 sAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThrThrGlnT 1121
3248 GCCTTACGCGCTGATTCGCGCGCGCGGCGTCCGTAAGACAGAA 3297
1121 hrGlySerSerIleThrSerSerAsn.....GlyGlnThrThr 1133
3298 AGCGTTCGCAACCGCGCGCGGACGCGGCGGAAAT..... 3336
1134 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaLysAsnVa 1150
3337GTGCGCATTTGACGCGCGGAGAGAAAGAAAGCGG 3373
1150 IThrLeuAsnThrThrGlyThrLeuThrThrThrGlyAspSerLys...I 1166
3374 TGCAGCGGATTAAGACACCGCTTGGCGAAGACAGCGAGCGGAAC 3423
1166 IeAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1182
3424 CGGCGGCTACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTT 3473
1183 AspGlyAlaAlaSerGly..... 1188

```

3474 GCGGCACTGCACACCCGACGCGCCGACGCGGACCTGATCA 3523
1189 .....AspArgThrValVala 1194
3524 GCCGTATGCCAATAGCGTTTGATGAAATTTCCGCCAGCTCAACAG 3573
1194 snalathrasnalaserGlySerGlyAsnValThrAlaLysThrSer 1210
3574 GTTTCCCGCTACAGACGAATTAAGC.....CGCGATT 3608
1211 SerValasnThrThrGlyAspLeuAsnThrThrIleasnGlyLeuasnIle 1227
3609 TCGCGAAGACCGCGCAACCGCGTTTGACACAGCGGATCCG...GACA 3655
1227 eserGlyasnGlyArgAsnThrValArgLeuArgGlyLysGluIleasp 1244
3656 CCAAACTACTACCGTTGCGAAGATTTCGCGCGCTCCGCAACAAACGAG 3705
1244 allystyryle.....GlnPro.GlyValAlaSerValGluValI 1258
3706 CTGCGCAATCGGTATGCAGAAAAC..... 3733
1258 eglualalysrrgvalleuglulysvalIlyasplseuSerAspGlu 1275
3734 .....TCGGAGCGCGCGCGCTCGCATCTCTTTTCGACACACGCGAGC 3778
1275 rggluthrleuAlaLysleuglyValSerAlaValArgPheValGluPro 1291
3779 AAAACACCTTCGACGCGCATCGGCACTCGCGCGGCTTCCCGACGCG 3828
1292 AsnAsnAlaIleThrVal.....AsnThrGlnAsnGluPheThrThr 1306
3829 GCCGTTTCGGGCAATAGCGCATGCACAGGTCTACA.....TCGCGAT 3872
1306 sprosSerGlnValThrIleSerGluGlyLysAlaCysPheSerSerg 1323
3873 CAGCGCGCGCGCGGTTTATAGCAGCGCGAGCC 3904
1323 IyAsnGlyAlaArgValCysThrAsnValAla 1333
seq_name: /cgn2_6/prodata/1/laa/6B_COMB.pep:us-08-719-641-9
seq_documentation_block:
; Sequence 9, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-9

```

```

alignment_scores:
Quality: 281.00 Length: 1478
Ratio: 0.404 Gaps: 72
Percent Similarity: 47.023 Percent Identity: 20.162

```

```

alignment_block:
US-09-303-518D-649 x US-08-719-641-9 ..

```

```

Align seg 1/1 to: US-08-719-641-9 from: 1 to: 1338

```

```

127 GCACACACTATTTTCGGCATCACTACCAATACATTCGCGACTTCCCGA 176
89 GlyGluThrThrLeuGly.....GlyAsn 96
177 AATAAAGCAAGTTTGCAGTCGGCGCGAAGATATGAGTTTACAACA 226
96 pGluArgGlyGlu.....GlyLysAsnGlyLeuGlnLeuAlaLysL 110
227 AAAAAGGGAGTTGGTCGCAATATGACAAAGCCCGATGTTTAT 276
110 ys.....ThrThrLeuGluLysGlySerThrIleAsn 120
277 TTTTCTGTGTCGCGATACGCGCGCGGCGCATGTGGCGCATCAATA 326
121 ValSerGlyLysGluLysGlyValArgAlaIleValITrpgLysPheAl 137
327 TATGTG.....AGCGTGCATTAACG 349
137 aleuIleAspGlyAsnIleAsnAlaGlnClyLysAspIleAlaLysThr 154
350 GCGGCTAT..... 357
154 IyGlyPheValGluThrSerGlyHisTyrlLeuSerIleAspAsnAla 170
357 ..... 357
171 IleValLysThrLysGluThrLeuLeuAspProGluAsnValThrIleG 187
358 .....AACACGTTGATTTTGGTGGAGAGGAAGAAATCCCG 394
187 uAlaProSerAlaSerArgValGlnleuGlyAlaAsp...ArgAsnSerH 203
395 ATCAACATGCTTTTACTTATAAATTTGCAACGGAATATATATAAGCA 444
203 IaserAlaGluValIleLysValIThrLeuLysAsnAsnThrSerLeu 219
445 GGGACTAAAGCCCATCTATGCGCGCATTTATCATATGCCGCTTTCGA 494
220 ThrThrLeuThrAsn.....ThrThrIleSe 228
495 TAAATTTGTACAGATGACAAACCTGTGAATAAGCAAGT..... 534
228 rAnleuLeuLysSerAlaHisValValAsnIleThrAlaArgAlaLysL 245

```

```

535 ..TATATGATGGGGGAAATATATGATCAAAATATATACCTGACCGT 582
      ::::::::::::::::::::
245 euThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleu 261
      ::::::::::::::::::::
563 GTTCTATTGGGGGAGGAGCAATATGGGATCTGATGAAGTAGACC 632
      ||| ||| ||| ::::::::::
262 HisSerGluGlyGlnGlyGlnGlyValGlnIleAspLysAsp.... 276
      ::::::::::::::::::::
633 CAATTAACCGCGAAAGTTCATATCATATGCAAGGCGTATTCTGGCTGC 682
      |||:::||||
277 .....IleThrSerGlu..... 280
      ::::::::::::::::::::
663 TTGGTGGCAATACCTTTCACAAAATGATCAGGTGT..... 720
      ||| ||| |||
281 ..GlyGlyAsnLeuThrIleTyr.....SerGlyGlyTyrValAspVal 294
      ::::::::::::::::::::
721 ..GGCAGATCACTTAGTAGTGAATAAAATTAACATACCCATATAG 767
      ::::::::::::::::::::
295 HisLysAsnIleThrLeuGlySer.....G1 303
      ::::::::::::::::::::
768 TTTTATACACAGAGGCTCATTTGGGAC.....A 799
      ||| ||| |||
303 yPheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLys 320
      ::::::::::::::::::::
800 GTGGCTCAACAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTTAAT 849
      ||| ||| |||
320 erGlyArgAsnAsnLeuThrIleThrAlaGln..... 330
      ::::::::::::::::::::
850 AATGGGATATGCAAGGGCAACCCCTATATAGAAAACATAGCTT 899
      |||:::||||
331 ..GlyThrIleThrSerGlyAsn.....SerAsnGlyPh 341
      ::::::::::::::::::::
900 CCAGCTGGTGTGTAACAGATGTTCTATGATGAATCTTGTGAGAGTA 949
      ||:|
341 eaArg.....PheAsnAsnValSerLeuAsnSerL 351
      ::::::::::::::::::::
950 CCCATTGATATTCTACAGACACGCTCAAAATGGAAATACTCTTTTAC 999
      ::| ||| |||
351 eu.....GlyGlyLysLeuSerPheThr 358
      ::::::::::::::::::::
1000 GACGATAT..... 1008
      |||:::||||
359 AspSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 375
      ::| ||| |||
1009 .....AATGGCAAGAAAATCAATGCAAC 1036
      ::| ||| |||
375 sPheAspLysThrLeuAsnIleSerGlyThrValAspIleSerMetLysA 392
      ::| ||| |||
1037 ATGAACACAAATTCCTGCTAATAGATTAAAGACGAAACGTT..... 1080
      ||| ||| |||
392 laProlValSerTyrPheTyrArgAspLysGlyArgThrTyrPasn 408
      ::| ||| |||
1081 .....CAATTGTTAATGTTCTTATAC 1103
      ::| ||| |||
409 ValThrThrLeuAsnValThrSerLysSerLysPheAsnLeuSerIleAs 425
      ::| ||| |||
1104 CGAGACACAGAGAAAGCTGTTTATCATCTGACAGGTGTCACAGTT 1153
      ::| ||| |||
425 pSerThr.....GlySerGlySerT 432
      ::| ||| |||
1154 ATCGACCCAGACTGATATATGAGAA.....ATATTTCTTT..... 1191
      ||| ::| ||| |||
432 hngLysProSerIleArgAsnAlaGlnLeuAsnGlyIleThrPheAsnLys 448
      ::| ||| |||
1192 .....ATTGACGAAGAAAGCGCAATGATGATCTTACAGCAA 1229
      ||| ::| ||| |||
449 AlaThrPheAsnIleAlaGlnLysSer.....ThrAlaAs 460
      ::| ||| |||
1230 CATCATCAAGGTCTGAGAGATTATTTCCAAAGAGATTTTACGGCT 1279
      |||:::||||
460 nPheSerIleLysAlaSerIleMetProPheLysSerAsnAlaSerLys 477
      ::| ||| |||
1280 CGCTGAAAAATACGAACCTTGGCAAGCGCGGCTTCATATACAGTGA 1329
      ::| ||| |||

```

```

477 laLeuPheAsnGluAsp.....IleSerValSerGly 487
      ::| ||| |||
1330 GACAGTACCGTTACTTGGAAAGTAAACGGC..... 1359
      ::| ||| |||
488 GlyLysSerValAsnPheLysLeuAsnAlaSerSerAsnIleGlnThr 504
      ::| ||| |||
1360 .....GTGGCAACGACCGCTGTCCAAATATGCGCAAGGCA 1396
      ::| ||| |||
504 rProGlyValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerT 521
      ::| ||| |||
1397 CGCTGACGTTCAAGCCAAAGGGGAAACCAAGCTCATACAGCTG... 1443
      |||:::||||
521 hIleuAsnLeuLysAlaGlnGlySerThrGlnThrAlaPheSerIleGlu 537
      ::| ||| |||
1444 .....GGGACGCTACACTCATTTTGGCA 1466
      ::| ||| |||
538 AsnAspLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnVa 554
      ::| ||| |||
1467 TCAGACAGCAGACGATPAA..... 1485
      ::| ||| |||
554 lGluGlyThrAspSerArgValAsnLysGlyValAlaAlaLysLysAsnI 571
      ::| ||| |||
1486 .....GGCAAAAACAGCCTTTAGT 1506
      |||:::||||
571 leThrPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThr 587
      ::| ||| |||
1507 GAATGCGCTTGGTACAGCGGAGGCTACGCTCAACTGATGCCATTA 1556
      ||| ||| |||
588 GluIle.....LysGlyAsnValThrIleAsnLysAsnTh 599
      ::| ||| |||
1557 T.....CAGTTCAACCCCGACAAACTCATTTTGC 1585
      ::| ||| |||
599 rAsnAlaThrLeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeuA 616
      ::| ||| |||
1586 GCTTTGCGGGGAGCGTTGGATTAAACGGGCAATTCGTTTCGCAC 1635
      ::| ||| |||
616 snIleAlaGlyAsnValIleAsn...AsnGlyAsnLeuThrAlaGly 631
      ::| ||| |||
1636 CGTATTCAAATACCGATGAGGGGCGATGTCATACCAATCAATAA 1685
      ||| ||| |||
632 SerIleIleAsnIle...AlaGlyAsnLeuThrValSerLysGlyAlaAs 647
      ::| ||| |||
1686 CAAAGATCCACCGTTACCATTTACAGCAATAAGATTATGCTACAACCG 1735
      ::| ||| |||
647 nLeuGlnAlaIleThrAsnTyrThrPheAsn.....ValAlaGlySerP 662
      ::| ||| |||
1736 GCATTAACAACACGCTTGATAGCAAAAAGAAATGCTTACACGCTGG 1785
      ::| ||| |||
662 heAspAsnAsnGlyAla...SerAsnIleSerIleAla...ArgGlyGly 676
      ::| ||| |||
1786 TTTCGGCAGAAAGATAGACCAAAAGACGGGCGCTCACTTCTTTA 1835
      ||| ||| |||
677 AlAlaLysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrTh 692
      ::| ||| |||
1836 CCAGCCCGCCGAGAAAGACCGCACCTGCTTTCGGGGAACAATT 1885
      ::| ||| |||
692 rAsnSerAspThrThrTyrThrGThrIle.....I 702
      ::| ||| |||
1886 TAAACGGCAATCAGCAACAACAACGCAAACTGTTTTCACGGCGACA 1935
      ::| ||| |||
702 leLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIleAspLys 718
      ::| ||| |||
1936 CCAACACCGCAGCGCTAACAATCATTTAAACGACCACTGTGCGCAAAA 1985
      ::| ||| |||
719 LysSer...AspAlaGluIleGlnIleGlyGlyAsnIleSerLysIle 734
      ::| ||| |||
1986 GGGCATTCCTCGGGGGAATCGTGTGGGACAACAGCTGGATCAACCGCA 2035
      |||:::||||
734 u.....GlyAsnLeuThrIleSerSerAspLysValAsnIleT 747
      ::| ||| |||
2036 CA.....TTTAAACGCGAAAC 2052
      ::| ||| |||

```

747 hrasnGlnIleThrIleLysAlaGlyValGlyIleArgSerAspSer 763
2053 TTCCAATTAAGGCGGACGCGGTGTTCCCGCAATGTTGCCAAGT 2102
::: :::::::::::::: :::::::::: :::
764 SerGlnAlaGlnAlaAsnLeuThrIleGlnThrLysGlnLeuLysLe 780
2103 GAAAGCGGATTTGGCATTTGACCAATCACGCCCAAGCAAGTTTGGTGTG 2152
: :::::::::::::: :::::::::: :::
780 uAlaGlyLysPheLysAsnIleSerGlyPheAsnLysAlaGlnLeuThrAlaL 797
2153 CACGCGCATCAAGCCACACATCTGTACACCTTGGACCTGGACGGGTCTG 2202
797 ysaSnGlySerAspLeuThrIleGlyAsnAlaSerGly..... 809
2203 ACAATTGTGTGAAAAAACCATTCAGCGACGATTAAGTATGCTTCAT 2252
::: :::::::::: :::::::::: :::
810 GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLys 826
2253 GACTAAGCGGACATCACGGGCAATGTGATCTTCCGATCAGCTCAT 2302
::: :::::::::: :::::::::: :::
826 sIleSerThrAsp...GlyHisAsnValThrLeuAsnSerGlu..... 839
2303 TAAATCTACAGGGGCTTGCACACTCAACGCGCATCTTAGTCAAAATGGC 2352
840ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 851
2353 GAT.....ACAGCTTATACAGTCAGC...CACACGCCACCCAAACG 2393
::: :::::::::: :::::::::: :::
852 AspAsnSerThrGlyLeuThrIleSerAlaLysAspAlaThrValAsnAs 868
2394 CAACCTTAGCTCGTGGGCAATGCCCAAGACATTTAATCAGCCACAT 2443
: :::::::::: :::::::::: :::
868 nasValAlaThr.....SerHisLysThrIleAsnIleSerAla 881
2444 TAAAGCGACACATCGGCTTGGGCAATGCTTCATTTAATCTAAGGAC 2493
881 laAlaGlyAsnValThrThrLysGlnGlyThrIleAsnAlaThrThr 897
2494 CACGCGCGTA.....CAAAAGCGCAGTCTGACGCTTCCGCGCA 2531
: :::::::::: :::::::::: :::
898 GlySerValGlnValThrAlaGlnAsnGly.....ThrIleLysGlyAs 912
2532 C...GCTAAGCAACGTAAGCCATTCGCGACTCAAGCGTAATGTCTCC 2578
: :::::::::: :::::::::: :::
912 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 928
2579 TAGCGGATAGGAGCATTCATCTTTGAAGC.....AGC 2613
: :::::::::: :::::::::: :::
928 hrThrGlnAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 944
2614 CGCTTTACGGACAATCAGCGGCGCAAGATAGGCA.....TT 2654
: :::::::::: :::::::::: :::
945 ThrLysThrGlyAspIleLysGlyIleGlnLeuThrSerGlyAsnVal 961
2655 ACACTTAAACAGACGCAATGACGCTGCCGTCAAGCAGCGAATTAGCA 2704
: :::::::::: :::::::::: :::
961 lAsnIleThrAlaSerGlyAsnThrLeu.....LysValSerA 974
2705 ATTTAACCTTGACAGCGCCACATTCACATCACTATCCGCTATCGCAC 2754
: :::::::::: :::::::::: :::
974 snIleThrGlyGlnAspAlaThrValThrAlaAsp..... 985
2755 GATCGGAGGAGGCGCAAAACGCGGATGCGACAGATGCGCGCGCGCG 2804
: :::::::::: :::::::::: :::
986AlaGlyAlaLeuThrThrAla..... 993
2805 TTCGCGCGGCTTCGCGCTTCCTATTATCGTTACACGCGCAACTTGG 2854
: :::::::::: :::::::::: :::
994GlySerThrIleSerAlaThrThrGlyAsnAla. 1004
2855 TAGAATCCCGTTTCAACACGCTGACGCTAAAGCGCAATTTGAACGCTGAC 2904
: :::::::::: :::::::::: :::
1005AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1016
2905 GGAACATTCGCTTATGTGGAACCTTCGGCTACCGAGCGCAAAAT 2954
1016 1016
2955 GAAGTCGCGGAAAGTTCGGAAGCACATTACCTTTGGCGGTCAACAT. 3003
: :::::::::: :::::::::: :::
1017ValGlnSerSerGlySerValThrLeuValAlaThrGlyA 1031
3004ACCGCAACGAACCTGCAAGCCTCGAACATTTGACGGA 3042
: :::::::::: :::::::::: :::
1031 laThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAla 1047
3043 GTGGAAGAAAAGACAAACCGCTGTCGAAAACCTTAAT..... 3084
: :::::::::: :::::::::: :::
1048 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1064
3085TTCACCCCTGCANA 3097
1064 nSerValThrThrSerSerGlnSerGlyAspIleGlnGlyThrIleSerG 1081
1081 LysAsnThrValAsnValThrAla.....SerThr 1090
3148 GCGGACTTCCGCTGCATATCCGGTCAAAAGACAGCTTTCGACAA 3197
: :::::::::: :::::::::: :::
1091 GlyAspLeuThrIleGlyAsnSerAlaLys.....ValGlnAlaLys 1104
3198 ACTCGGCAAGCGCAGAAAGCCAAAAACAGCGGCAAAAGCAAGCGGCAAA 3247
: :::::::::: :::::::::: :::
1104 saSnGlyAlaAlaThrLeuThrAlaGlnSerGlyLysLeuThrGlnT 1121
3248 GCCTTACGCGCTGATTCGCGCGGCGGATGCGTCAAAAGACAA 3297
: :::::::::: :::::::::: :::
1121 hrGlySerSerIleThrSerSerAsn.....GlyGlnThrThr 1133
3298 AGCGTTCCGCAACCGGCGCGGACGAGCGGGGAAAT..... 3336
: :::::::::: :::::::::: :::
1134 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnVal 1150
3337GTCCGATTTATGACGCGGAGGAGCAAGAAAGCAAGCGG 3373
: :::::::::: :::::::::: :::
1150 lThrLeuAsnThrThrGlyThrLeuThrThrLysAspSerLys...I 1166
1166 lAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1182
3424 CGGCGGCTACCAACGCGCTTCCCGCGCGCGCGCGCGCGCGGAGATT 3473
: :::::::::: :::::::::: :::
1183 AspGlyAlaAlaIleSerLys..... 1188
3474 GCGGCAACTGCAACCCCAACCGGACGCCCAACGCGGCGGCGCGATCA 3523
1189AspAlaThrValValA 1194
3524 GCGGTTATGCAATAGCGGTTTGAGTAATTTCCGCAAGCTCAACAGC 3573
: :::::::::: :::::::::: :::
1194 snAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSer 1210
3574 GTTTTCGCGCTACAGACGAATTAGAC.....CGCGTAT 3608
: :::::::::: :::::::::: :::
1211 serValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleI 1227
3609 TCGCGAAGACCGCGCAACGCGCTTGGACAAGCGGATCCGG...GACA 3655
: :::::::::: :::::::::: :::
1227 eSerGlnAsnGlyArgAsnThrValAlaArgLeuArgGlyLysGlnIleAspV 1244
3656 CCAAACTACTACGCTGCGCAAGATTCGCGGCTTACGCGCAACAAACGAC 3705
: :::::::::: :::::::::: :::
1244 alLysThrIle.....GlnPro.GlyValAlaSerValGlnGlnValAl 1258


```

800 GGGGCTCAACAATGTTTATTCATGATGAGCCCAAGCAAAAGCTGTTATT 849
      |||||      :::      |||||
580 erglYarqasnasnleuthrllethralagln..... 590
850 AATGGGGAATTGCAACGGGCAACCCCTATTATGAGAAAAACAATGGCTT 899
      |||||      :::      |||||
591 ...gltYthrlllethrserglYasn.....seranglysn 601
900 CCAGCTGGTCTGAAGATGGTTCATGATGAAATCTTTGTGGAGATA 949
      |||||      :::      |||||
601 earg.....Pheasnasvalseusnserl 611
950 CCCATTACGATTATTCAGCAACACGCTCAAAATGCAATACTCTTTTAAC 999
      |||||      :::      |||||
611 eu.....glyglYsleuserPheThr 618
1000 GACGATAT..... 1008
619 AAsperaragluAsparglYarqarYthrlysglYasnllseasnll 635
      |||||      :::      |||||
1009 .....AATGGCAGAGAAAAATCAATGCCCAAC 1036
635 sPheaspglYthrleuasnlleserglYthvalAsplllesermetlysa 652
1037 ATGAACACAATTCTCTGCTATATGATTAAAAACAGACCGTT..... 1080
      |||      |||      |||||
652 lAprolyvalserlrrPheYrargAspLyglYarqThYrTrrpsn 668
1081 .....CAATTGTTAATGTTCTTATAC 1103
669 valThrThleuasnvalThrseryserlyspheasnleuserlleas 685
1104 CGAGACAGCAAGAGAACCTGTTTATCATGCTCGAGGTGCTCAACAGTT 1153
      ::|||      |||      ::|||
685 pserThr.....glyserglyser 692
1154 ATCGACCCAGACTGAAATATGGAAGA.....AATATTTCCTTT..... 1191
      |||      ::|||      |||||
692 hrclyproserlleargnasnlaagluleuasnglyllethrpheasnlys 708
1192 .....ATTGACAGAGAAAGCAATGATACTACTACAGCA 1229
709 Alathrpheasnllleaglndlyser.....Thralaas 720
1230 CATCAATCAAGCTGCTCGAGCAATTATATTCACAAGAGATTTACGCTCT 1279
      |||      ::|||      ::|||
720 nPheSerllelysaIaserllemeProphelyserasnlleasnlyra 737
1280 CGGCTGAATAATPACGAACCTTGCAAGCGCGGGGCTTCATATAGGAA 1329
      ::|||      ::|||
737 laeupheasnclusp.....llesevalsergly 747
1330 GACATACCGTTACTTACGAAGTAAACGGC..... 1359
      ::|||      ::|||      ::|||
748 glyglYservalAsnphelyseusnaslaserSerasnllleglnTh 764
1360 .....GTCGCAACAGACCGGCTGCCAAATTCGCAAGCA 1396
764 rProglyvallllelleysserglAsnphesnasnalseerlylyserT 781
1397 CGCTGACGCTTCAAGCAAAAGGGAACAACAAGCTGCATACGCGTG... 1443
      |||||      |||||      |||||
781 hrlleuasnleuysnlaaglYserThrngluthAlaPheSerlleglu 797
1444 .....GGCGACGCTACAGCTATTGGGA 1466
798 AsnAspLeuasnleuasnAlaThrglyglYasnllleThrlllearglnva 814
1467 TCAGCAGGAGAGCGATTA..... 1485
814 lgluglYthrrAspserarvalAsnlysglYalAlaAlaYslYasnT 831

```

[illegible]

```
1100 .....ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 1111
2353 GAT.....ACACGTTATACAGTACAG...CACACGGCACCCCAACAGG 2393
1112 AspAsnSerThrGlyLeuThrIleSerAlaIAspValThrValAsnAs 1128
2394 CAACCTTACCTCTGTGGGCAATGCCAGCAACATTATATCAAGCCAGAT 2443
1128 AsnValThr.....SerHisLysThrIleAsnIleSerAla 1141
2444 TAAAGGCAACACATCGCTTGGGCAATGCTTCAATTATATTAAGCAGC 2493
1141 AlaIAsnValThrThrLysGlyThrIleAsnAlaThrThr 1157
2494 CACGCGTA.....CAAAAGCAGCTGACGCTTCCGGGCA 2531
1156 GlySerValGluValThrAlaGlnsngly.....ThrIleLysGlyAs 1172
2532 C...GCTAAGCAACGTAAGCCATTCCGCACTCAACGGTATGTCTCC 2578
1172 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 1188
2579 TAGCGGATAGGCACTATTCATTGGAAGC.....AGC 2613
1188 hThrGlnAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1204
2614 CGCTTACCGCAATACAGCGCGCAAGATACGGCA.....TT 2654
1205 ThrLysThrGlyAspIleLysGlyIleGlnSerThrSerGlyAsnVa 1221
2655 ACACCTTAAAGACGCAATGACGCTGCGCTCAGCAGCAGGAATTAGCA 2704
1221 AsnIleThrAlaSerGlyAsnThrLeu.....LysValSerA 1234
2705 ATTAAACCTTGACACGCCACCATTAACATTCATTCGGCTATCGCAC 2754
1234 snIleThrGlyGlnAspValThrValThrAlaAsp..... 1245
2755 GATCGGCGAGGCGCAACCGGCGAGTCGCGAGATCGCGCGCGCG 2804
1246 .....AlaGlyAlaLeuThrThrAla..... 1253
2805 TTCGCGCGCTTCGCGCTTCCATTAATTCGCTTACACCGCAACTTCG 2854
1254 .....GlySerThrIleSerAlaThrThrGlyAsnAla 1264
2855 TAGATCCCGCTTTCACACGCTGACGTAACGGCAATTAAGAGGTGAG 2904
1265 .....AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1276
2905 GGAAACATTCGCTTATGTGGAATCTTCGCTACCGCAGCAGCAAAAT 2954
1276 ..... 1276
2955 GAAGCTGGCGGAAATTCGGAAGCACTTAACCTGGCGGTCAACAT. 3003
1277 .....ValGlnSerSerGlySerValThrLeuValAlaThrGlyA 1291
3004 .....ACCGCAACGACACTGCAAGCTGCAACATTAAGGTA 3042
1291 laThrLeuAlaValIleAsnIleSerGlyAsnThrValThrIleThrAla 1307
3043 GTGGAAGAAAAAGACAAACCCGTGTCCGAAAACTTAAT..... 3084
1308 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1324
3085 .....TTCACCTGCAGAA 3097
1324 nSerValThrThrSerSerGlnSerGlyAspIleGlnGlyThrIleSerG 1341
3098 ACGAACACGTGATGCGCGCTGTACCAACTCATCCGCAAGAC 3147
..... 3147

1341 LysAnThrValAsnValThrAla.....SerThr 1350
3148 GGGCAGTTCCGCCCTGCATTAATCCGCTCAAGAACAGACTTCCGACAA 3197
1351 GlyAspLeuThrIleGlyAsnSerAlaLys.....ValGluAlaLys 1364
3198 ACTCGCAAGCAGACAGCCAAACAGCGGAGAAAAAGACACGCCGCAA 3247
1364 AsnGlyAlaLeuThrLeuThrAlaGlnSerGlyLysLeuThrThrIle 1381
3248 GCCTTGACGCGCTGATTCGCGCGCGCGCGATGCCGTGAAAAGACGAA 3297
1381 hGlySerSerIleThrSerSerAsn.....GlyGlnThrThr 1393
3298 AGCGTGCAGCAACCGCGCGCGCGCGCGGCAAAAT..... 3336
1394 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaIAsnVa 1410
3337 .....GTCGCATTATGACGCGGAGAGAGAGAAAAACGCG 3373
1410 lThrLeuAsnThrThrGlyThrLeuThrThrThrGlyAspSerLys...I 1426
3374 TGCAGCGCGATTAAGACACCGCCTTGCGGAACACGCGGAGGAAACC 3423
1426 leAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1442
3424 CGGCGGCTACCAACCGCCTTCCCGCGCGCGCGCGCGCGGATTT 3473
1443 AspGlyAlaIleAsnSerGly..... 1448
3474 GCGCAACTGCACACCCCAACCGCAGCCCAACCGCAGCGGACCTGATCA 3523
1449 .....AspArgThrValValAla 1454
3524 GCCGTTATGCCATAGCGGTTTGAGTGAATTTCCGCCACGCTCAACAG 3573
1454 snAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSer 1470
3574 GTTTTCGCCGTACAGGAGCAATTAAGAC.....CGCGTATT 3608
1471 SerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleI 1487
3609 TGCGGAAGACCGCGCAACCGCGCTTGGACAGCGGCATCCGG...GACA 3655
1487 eSerGlnAsnGlyArgAsnThrValArgLeuArgGlyLysGluIleAsp 1504
3656 CCAACACTACCGCTTCGCAAGATTTCCGCGCTACCGCACAACACCGAC 3705
1504 allysTyrlle.....GlnPro.GlyValAlaIleSerValGluGluValI 1518
3706 CTGCGGCCAAATCGTATGCAAAAAAC..... 3733
1518 eGluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluGln 1535
3734 .....TCGCAAGCGGCGCGCTCGCATCTGTTTGCACAAACCGGACG 3778
1535 rgGluThrLeuAlaLysLeuGlyValSerAlaValAlaArgPheValGluPro 1551
3779 AAAACACCTTCGACAGCGGATCGCAACTCGGACAGCGCTTGCCACGCG 3828
1552 AsnAsnAlaIleThrVal.....AsnThrGlnAsnGluPheThrThr 1566
3829 GCGGTTTTCGGCAATACGCAATCGACAGCTGATCTTAC.....TCGGCAT 3872
1566 sProSerSerGlnValThrIleSerGlnGlyLysAlaLysPheSerSerC 1583
3873 CACGCGGCGCGGCTTTAGCAGCGGACCC 3904
1583 LysnGlyAlaArgValLysThrAsnValAla 1593

seq_name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:US-08-038-682-4
seq_documentation_block:
```

```
; Sequence 4, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; SUITE: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0266
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERSSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-4

alignment_scores:
  Quality: 282.50      Length: 1388
  Ratio: 0.428         Gaps: 61
  Percent Similarity: 47.550   Percent Identity: 19.380

alignment_block:
US-09-303-518D-649 x US-08-038-682-4 ..

Align seg 1/1 to: US-08-038-682-4 from: 1 to: 1477

175 GAAATATAAGCAGATTGTCAGTGGGCGAAGATATGAGTTTACAA 224
:::|||||:::|||||:::|||||:::|||||:::|||||:::
219 Lysasnuglyvalilleservalasnlylyserilleserleu... 233
225 CAAAAAGGAGTGTGGCGCAATCAATGCAAAAAGCC... 264
|||||:::|||||:::|||||:::|||||:::|||||:::
224 .....LeuAlaglylnlylserilseraspillerlea 246
265 ..CCGATGATGATTTTCTGTGTCGCGTAAAGCGCGGCGCATG 312
||| ||| ::|||:::|||||:::|||||:::|||||:::
246 snprrhrllethrrtyserllealaIaproguaangluAlaValasn 262
313 GTGGCGCATCAATATTTGTAGCGTGCACATTAAGCGGCGCTATACAA 362
:::|||||:::|||||:::|||||:::|||||:::
263 leuglyasp.....llehealalysgllyasnllleas 274
363 CGTT.....GATTTGTCGCGGAGAA 385
||| ||| ::|||:::|||||:::|||||:::|||||:::
274 nvalalrgalalathrlleatrgasnnglylyleuSerAlasperv 291
386 GAAATCCGATCAACATCGTTTACTATATAAATTGAAAGGAAATTAAT 435
```

```
::: |||::: ::|||:::
291 alserlyaspyserserlyasnlllevalleuSerAlalysgllyglu 307
436 TATAAGCAGG.....ACTAAAGCCATCCTTATGGCGG 470
::: |||::: ::|||:::
308 Alagluilleglyvalilleseralaglnsnnglnalalalysgllygl 324
471 CGATTATCATATGCGCGTGTGCATTAATTTGCACAGATCAGACACCTG 520
| ::: ::: :::
324 lysleuethrllethrglyaspysvalthrleuThrlythrlyAlaValI 341
521 TTGAATATGACAGTTATATGATGGCGGGAATATATGATCAAAATAT 570
::: |||::: |||::: |||::: |||:::
341 leasplesercllysglnlygllygluthrlyleuuglyaspelnu 357
571 TACCTGACCGGTGTCATGTCGAGGACGACGCAATATGTCGATCGA 620
358 Arg.....glyluuyllyasnnglyllelnleuAl 368
621 TGAGATGAGCCCAATTAACCGGAAAGTTCATATCATATTCAGAGT... 666
::: |||::: |||::: |||::: |||:::
368 alyslsthrserleuaglyllysglyserthrilleasnvalserglysg 385
667 .....CGCTATTCCTGC.....CTGGTGGTGGC 690
385 llysglylypheAlalvaltrpglyaspillealaleuileaspely 401
691 AATACCTTTGCAAAATGATGATGAGTGGTGCACAGTCAACTTAGTATG 740
||||| ||||| ||||| |||||
402 AsnllleasnAlagln...glyserly..... 409
741 TGAATAATTAACATATAGCCCATATGTTTATACCAACAGAGGCTCAT 790
410 .....Aspillealalystrlyglylyphev 418
791 TTGGGACAGTGCCTCACCAATGTTATCTATGATGCCAAAGCAAAAG 840
||||| ::||| |||
418 algluthsercllyhisaspheupheilleysasp..... 429
841 TGGTTAATTAATGGGATATGCAAAAGGCAACCCCTATATAGCAAAAG 890
|||||:::|||||:::
430 .....AsnAlalilleAlaspAla..... 435
891 CAATGCTCCAGCTGCTGCTGAAGATGCTTCATGATGAATCTTTG 940
436 .....Lysgluthrpleuenusp..... 441
941 CTGAGATACCCATTCAGTATCTACGAACGCTCAAAATGGGAATATC 990
|||::: |||::: |||
442 ..PheaspasnvalserilleasnAlagluaspProleupheasnlnthr 457
991 TCTTTTAACGAGAT...AATATGACACAGGAAATCATATGCCAACA 1037
::: |||::: |||::: |||::: |||:::
458 GlylleasnaspglupheProthrlythrlygluAlaseraspProly 474
1038 TGAACCAATCTCTGCTCAATATAGATTAATAACAGCAACCGCTAATGT 1087
::: |||::: |||::: |||::: |||:::
474 slyasnserglleuLysThrlythrleuThrlythrlylleSerAsnt 491
1088 TT.....AATGTTCTTATTCGAGACAGACAGACA 1116
::: |||::: |||::: |||:::
491 yrlleuLysasnAlatrphrMetasnlllethrlAlaserArglyserthr 507
1117 GAACGTGTTATCATGCTGACAGGTGTGTCACAGCTTATCGACCCAGACT 1166
||| ||| ||| ||| ||| |||
508 .....Valasnser.....Serll 512
1167 GATTAATGACAAATATTTCTTATTCAGCAAGAAAGCGCAATGTA 1216
||| ||| ||| ||| ||| |||
512 easnllleglyserasnserhllleuilehllserlysgly..... 526
1217 TACTTACGACCAACATCAAGGTGCTGAGAGATTTATTTCCACAGA 1266
||| ::||| |||::: ::||| |||
```

```
527 .....GlnArglyglyValGlnIleAspGly 536
1267 GATTTACGGCTCGCCTGAAATAACGAACCTTGCGACGGCGGCGCT 1316
      ||| ||| |||
537 AspIleThrSerIleGlyAsnLeuThrIleTyrSerGlyGlyTyrPhe 553
1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAACGGCGTGGCAA 1366
      ||| ||| |||
553 AspValHisIleAsnIleThrLeu..... 561
1367 ACACCGCGCTGCCAAATCGGCAAGCAGCGCTGACGCTTACGCCAAA 1416
561 ..... 561
1417 GGGGAAACCAAGCTCGATCAGCGTGGCGACGGTACGATCTTTGSA 1466
      ||| ||| |||
562 .....AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe.. 575
1467 TCACGCGGCAAGCAGATTAAGCAAAAACAGCCCTTACGTAAGTCGGCT 1516
      ||| ||| |||
576 GlnGlyGlyAsnAsnIleAlaArgAspAlaAlaAsnAlaIle.. 590
1517 TGGTACGGCGAGGGGTACGGTCAACTGAATGCGGATATCAGTTCAAC 1566
591 .....ValAlaGlnGlyThrValThrIleThrGlyGlnGlyLys..... 603
1567 CCCGACCAACTCTATTTCGGCTTTCGGCGGCGACGCTTGGATTAAACGG 1616
      ||| ||| |||
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGln 614
1617 GCATTCGCTTCTGCTCCACCGTATTCAAAATCCGATGAAGGGCGATGA 1666
614 Y.....ThrGlyLysGlyLeuAsnI 621
1667 TTGTC.....AACCAATCAAGACAAAGAATCCACGCTTACC 1704
      ||| ||| |||
621 IleIleSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
1705 ATTACAGCAATTAAGATATTGCTCAACCGGCAATTAACACAGCTTGA 1754
638 IleSerGlyAsnIleThrIleAsnGlnThr..... 647
1755 TTAGCAAAAAGAATTCCTACACAGGTTGTTGGCGAAGATACG. 1803
648 ThrThrGlyAsnThrSerTyr.....TyrGlnThrSerHisAspSerH 662
1804 .....ACCAAAACGACGCGCGGCTCAACCTT 1830
662 LeTyrAsnValSerAlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPhe 678
1831 GTTTACACGCCCGCGCAGAGACCGCACCTGCTGCTT..... 1869
679 IleIleTyrIleSerSerAsnSerIleTyrThrGlnTyrArgSer 695
1870 TCCGCGGCAACAATTTAAAGCAACATCAACGAAACAAACGCGCAAC 1918
      ||| ||| |||
695 rSerIleGlyValAsnPheAsnGly.....ValAsnGlyAsnM 708
1919 TGTTTTTCAGCGCAGACCAACACCGCAGCGCTACATCATTTAAACGAC 1968
708 eTserPhe..... 710
1969 CATTTGGTGGCAAAAAGAGGCACTTCTCGCGGGAATCGTGTGGACAA 2018
711 .....AsnLeuLysGlnGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGATCAACCGCACCA.....TTTAAAGCGGAA 2050
725 nGlnAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla..A 741
2051 ACTTCCAAATTAAGCGGACGCGGCTGTT..... 2082
741 snIleThrAlaThrGlyGlySerValPhePheAspIleTyrAlaAsn 757

2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCAATTGACAA 2126
758 HisSerGlyArgGlyAlaGlnLeuLysMetSerIleIleAsnIleSerAs 774
2127 TCAGCGCCACAGCAGTTTGTGTGCGACCGCATCAAAAGCAGCACATCT 2176
774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783
2177 GTACAGTTTCGAGCTGACGCGGTCTGACAAATTTGTGCAAAAAACCAT 2226
784 ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATAAGTATGCTTCACTTGAACCCGACATCAGCGGCAA 2276
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTyr 815
2277 TGTGATCTTGGCGATCAGCTCATTTAAATCTCAGCAGCGCTTGCACAC 2326
815 rAspGlyTyrAlaLarGlyAsnAlaIleAsnSerThrTyrAsnIleSerIle 832
2327 TCAACGGCAATCTTATGCAAAATGCGATACGCTTATACAGTCAGCCAC 2376
832 euGlyGlyAsnValThrLeuGlyGly..... 840
2377 AACGCCACCCCAAGCGCAACTTACGCTGCTGGGCAATGCGCAAGCAC 2426
841 .....GlnAsnSerSerSerIleThrGlyLys.....IleThr 852
2427 ATTTAATCAAGCC.....ACATTAAACGGCAACATCGGCTTCGG 2467
852 rIleGlnLysAlaAlaAsnValThrLeuGlnAlaAsnAlaIleProAsn 869
2468 GCATGCTTCAATTAATCAAGCAGCAGCCGCTACAAAACGGCAGCTCTG 2517
869 nGln.....AsnIleArgAspArgValIleIleLysLeuGlySerLeu 882
2518 ACGCTTCCGCGACGCTAAGGCAACGTAAGCATTCGCCACATCAGCG 2567
883 LeuValAsnGlySerLeuSerLeuThrGlyGlnAsnAlaIleLysGln 899
2568 TAATGCTCCCTACCGCATTAAGCGCAGTATTCAT.....TTGAAA 2608
899 yAsnLeuThrIleSerGlnSerAlaThrPheLysGlyLysThrArgAsp 916
2609 GCACCGCTTACCGCAATAATCAGCGCGGCGCAAGATACGATACAC 2658
916 nLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGlnIleAsn 932
2659 TTAAGACAGACGCAATGAGCGTCCGCTCAGGACGAGCAATTAAGCAATT 2708
933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAsp 949
2709 AAACCTTGACACAGCCACCATTTACACTCAATTCGCGCTATGCCACGAT 2758
949 uAsnIleThrThr..... 953
2759 CGGACGGGGCGCAAAACCGGACAGTGCAGATGGCGCGCGCGCTGCG 2808
954 .....HisAla 955
2809 CGCGCTTCGGCGCGCTTCCATTATTCGCT...TACACGCCCACTTCGG 2854
956 LysArgAsnGlnArgSerIleIleGlyAlaAspIleIleAsnLysLysG 972
2855 TAGAATCCGCTTCAACACGCTGACGTAACGCAAAATGAACGGCTAC 2904
972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGlnIleGlnIleGly 989
2905 GAACATTTCCGCTTATGTCGAACTTCTGGCTACGCGACGCAAAAT 2954
989 LysAsnIleSerGlnLysGlnGlyAsnLeuThrIleSerSerAspLysI 1005
```

```

2955 GAAGCTGGCGAAGTTCCGAGGCACTTACCTTGGCGGTCAACATA 3004
      ::::::::::
1005 eAnIleThr..... 1008
3005 CCGGCAACGAACCTGCAAGCCTGCACAAATTGACGGTAGTGAGAGAAA 3054
      ::::::::::
1009 .....LysGlnIleThrIleTylsGlyLe 1017
3055 GACACAAACCCGCTGCC.....GAAACCTTAATTTCACCC 3092
      ::::::::::
1018 AspElYgluAspSerSerAspAlaThrSerAlaIAsnLeuThrI 1034
      ::::::::::
3093 GCAAAAGCAACAGCTCGAT.....GCCGGCGCGT 3121
      ::::::::::
1034 elYsThrIlyeGluldeuIlyleuThrgluAspLeuSerIleSerGlyPhe 1051
      ::::::::::
3122 GCGCTTACCAATCATCCGCAAAAGCGC...GAGTTCCGCTGCATTAAT 3168
      ::::::::::
1051 snIysAlaGluIleThrAlaIysAspGlyAspLeuThrIleGlyAsn 1067
      ::::::::::
3169 CCGGTCAAAAGCAAGCTTCCGCAAACTCGCGCAAGCGACAGCAACCA 3218
      ::::::::::
1068 Ser.....AsnAspGlyAsnSerGlyAlaGluAla.. 1077
      ::::::::::
3219 AAACAGCGCGGAAAAAGCAACAGCGCAAGCCTTGACGCGTGAATGCGG 3268
      ::::::::::
1078 .LysThrValThrPheAsnAsnValys.....AspSerLysIleSerA 1092
      ::::::::::
3269 CCGGGCGCGATCCGCGTGAAGACAGCAAGCCTTGCGGACCGCGCGG 3318
      ::::::::::
1092 laAspGlyHisAsnValThrLeuAsnSerLysValIlyThrSerSer 1108
      ::::::::::
3319 CAGGACGCGCGGAAATGCGGATTAATGACGCGGAGAGAGCAAAAA 3368
      ::::::::::
1109 AsnIlyGlyArgGluSer..... 1114
      ::::::::::
3369 ACGGGTACGCGGATTAAGACACCGCCTTGCGGAAACAGCGGAGACGG 3418
      ::::::::::
1115 .....AsnSerAspAsnAspThrGlyLeu..... 1122
      ::::::::::
3419 AAACCCGCGCGGTACACCGCCTTCCCGCGCGCGCGCGCGCGG 3468
      ::::::::::
1123 .....ThrIleThrAlaIysAsnValGluValAlaIys 1133
      ::::::::::
3469 GATTTCGCGCAATTCGCAACCCGACCGCGCGCGCGCGCGCGG 3518
      ::::::::::
1134 AspIleThrSerLeu..... 1138
      ::::::::::
3519 GATCAGCGGTATGCAATAGCGGTTGAGTAATTTCCGCCACGCTCA 3568
      ::::::::::
1139 .....LysThrValAsnIleThrAlaSerGluIlyValThrThr 1152
      ::::::::::
3569 ACACCGTTTCGCGGTACAGAGCAATTAGCGCGTATTTGCCGAGAC 3618
      ::::::::::
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrIly 1169
      ::::::::::
3619 CGCGCAACGCGCTTTGGACAGCGGATCGGAGACCAACACTACCG 3668
      ::::::::::
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186
      ::::::::::
3669 TTCCG...ACATTTCCGCGCGCTACCGCAACAAACGACCTGCGCAAA 3715
      ::::::::::
1186 laThrValAspLeuThrThrIlySerGlySerIleGluAlaLysSer 1202
      ::::::::::
3716 TCGGTATGCAAAAAACCTGCGAGCGGCGCGTCCGATCCCTTTTCG 3765
      ::::::::::
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyIly..... 1216
      ::::::::::
3766 CACAACGCGACGAAACACCTTCGACAGCGCATCGCA..... 3805
      ::::::::::
1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1233
      ::::::::::
3806 .....ACTCGGCAACGCGCTTCCCGACGCGG 3829

```

```

1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249
      ::::::::::
3830 CCGTTTCGGGCAATACGCGCATCGACA.....GGTCTACATCGGCATC 3873
      ::::::::::
1250 ThrAlaThrGlyAsnThrLeuThrIleThrGluAlaGlySerIleIleIle 1266
      ::::::::::
3874 AGCGGCGCGCGGCTTTAGCAGCGGACGCTTTCAGACGCGCATCGAGG 3923
      ::::::::::
1266 rThrIlyeGlyGlnValAspLeuLeuAla.....GlnAsnGlySerIleA 1281
      ::::::::::
3924 CAAATCCGCGCGCGCTGCTGCTGCTATTCGCGCATTCAGCAGCATACCGG 3973
      ::::::::::
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297
      ::::::::::
3974 CCGGTTTCGGGATTCGCGCATCGCAACCGCATCGCGCAACCGCTAT 4023
      ::::::::::
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValI 1314
      ::::::::::
4024 TTCGTCCAAA 4033
      ::::::::::
1314 eAnAlaIys 1317
      ::::::::::

```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4

seq_documentation_block:

Sequence 4, Application US/08302832

Patent No. 5603938

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

OF INVENTION: of No. 5603938-Typeable Haemophilus

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/302,832

FILING DATE: 16-SEP-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US pct/us93/02166

FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-404

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1477 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-302-832-4

alignment_scores:

..... 453

038 LieserGlyashnIleIrrIleasngInIhr..... 647


```

3469 GATTTCGCCCACTGCACACCCGACGCCCAACCGCAGCGGACCT 3518
1134 AspIleThrSerLeu..... 1138
3519 GATCACCCTTATGCAATAGCGGTTTGATTAATTTCCGCCAGCTCA 3568
1139 .....LysThrValAsnIleThrAlaSerGluValThrThr 1152
3569 ACAGCGTTTCGCCGTACAGACGATTAAGCGGTATTTCGCCAAGAC 3618
1153 AlaGlySerThrIleAsnAlaThrSngIyAlaSerIleThrThyL 1169
3619 CGCCGCAACCGCGTTTGACAGCGGATCCGGACCAACCACTACG 3668
1169 sThrGlySplSerGlyThrIleSerGlyAsnThrValSerValSer 1186
3669 TTCGC...AAGATTCCCGCGCTACCGCCCAACCAACCTCGCCCAA 3715
1186 lArhValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202
3716 TCGGTATCGAAGAAACCTCGGACGCGGCGGATCGCTGTTTCG 3765
1203 GlyGluAlaAsnValThrSerAlaThrIleGlyGly..... 1216
3766 CACAAACCGGACGAAACACCTTCGACGCGGATCGCA..... 3805
1217 ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1233
3806 .....ACTCGGACGCGCTTCGCCACGCGG 3829
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249
3830 CCGTTTCGGCAATACGCGATCGACA.....GGTTCTACATCGGATC 3873
1250 ThrAlaThrGlyAsnThrLeuThrGluAlaGlySerSerIleThr 1266
3874 AGCGGCGCGCGGCTTTAGACGCGGACCTTCGACGCGGATCGGAG 3923
1266 ThrLysGlyIleValAspLeuLeuAla.....GlnAsnGlySerIle 1281
3924 CAATATCGCGCGCGGTGCTGATTAAGCGGATTCGACGACGATGCGG 3973
1281 lAglySerIleAsnAlaAlaAsnValThrLeuAsnThrIleGlyThrLeu 1297
3974 CCGGTTTCGGCGGATTCGCGATCGACCGGACATCGCGGACGCTAT 4023
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1314
4024 TTCGTCCAA 4033
1314 eAsnAlaLys 1317

```

seq_name: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-530-198-4

seq_documentation_block:

Sequence 4, Application US/08530198

Patent No. 5869065

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J

APPLICANT: ST. GENE III, JOSEPH W

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Maltare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERTSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-4

```

```

alignment_scores:
Quality: 282.50 Length: 1388
Ratio: 0.428 Gaps: 61
Percent Similarity: 47.550 Percent Identity: 19.380

```

alignment_block:

US-09-303-518D-649 x US-08-530-198-4

```

Align seg 1/1 to: US-08-530-198-4 from: 1 to: 1477
175 GAAATTAAGCAAGTTTGCAAGTGGGCGCAAGATATTGAGGTTACAA 224
219 LysAsnGluValIleSerValAsnGlySerIleSerLeu..... 233
225 CAAAAAAGGAGTGTGCGCAATCAATGACAAAAGCC..... 264
234 .....LeuAlaGlyIleLysIleThrIleSerAspIleIle 246
265 ..CCGATGATTCATTTCGTGTCGTGTCGTCGTCGTCGTCGTCGTCG 312
246 snProThrIleThrIleThrIleAlaIleProGluAsnGluAlaValAsn 262
313 GTGGGCGATCAATATATTTGTGAGCGTGACATTAACGGCGGCTATAAC 362
263 LeuGlyAsp.....IlePheAlaLysGlyAlaAsnIleAs 274
363 CGTT.....GATTTCGTCGGGAAGCA 385
274 nValAlaGluAlaIleThrIleArgAsnGluLysLeuSerAlaAspServ 291
386 GAAATCCGATCAACATCGTTTACTTATTAATTTGTGAACGGAATAT 435
291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluVal 307
436 TATAAGCAGCG.....ACTAAGCCATCTTATGGCGG 470
308 AlaGluIleGlyValIleSerAlaGluAsnGluAlaLysGly 324
471 CGATTATCATATGCGCGGTTTCATTAATTTGTACAGATGAGAACTG 520
324 LysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaVal 341
521 TTGAATGACCAAGTTATATGATGGCGGAAATATATGATCAATATAT 570
341 LeuAspLeuSerGlyLysGluGlyGlyIleThrLeuGlyLysAspLeu 357
571 TACCTGACCGGTTCGATTGGGCGGACGAGCAATTTGCGGATCTGA 620

```

```
358 Arg.....GlyGluGlyLysAsnGlyIleGlnLeuAl 368
621 TGAAGATGAGCCCAATTAACCGGAAAGTTCAATATATTCGAAGT... 666
    ::::: ||||| ::::: |||||
368 alyslsYThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385
667 .....GGCTATTCCTGG.....CTCCTGGGTGGC 690
    ||| ||| ||| ||| ||| |||
385 lylsGlyGlyPheAlaIleValTTPGlyAspIleAlaLeuIleAspGly 401
691 AATACCTTTGCACAAATAGATCAGGTGGTGACAGTCACACTTAGTAG 740
    ||| ||||| ||||| |||||
402 AsnIleAsnIleGln...GlySerGly..... 409
741 TGAATAATTAACATATAGCCCATATGCTTTTACCACAGAGGCTCAT 790
    ::| ||||| ||||| |||||
410 .....AspIleAlaYThrGlyGlyPheV 418
791 TTGGCAGAGTGGCTCAGCAATGTTATCTATGATGCCCAAAAGCAAG 840
    ||||| ||||| ||||| |||||
418 alylulThrSerGlyHisAspLeuPheIleLysAsp..... 429
841 TGGTTAATTAATGGGTATTTGCAACGGCAACCCCTATATAGAAAAAG 890
    ||| ::::: ||::: ||:::
430 .....AsnAlaIleValAspAla..... 435
891 CAATGGCTCCAGCTGGTTCGTAAGATTTGTTCTATGATGAATCTTGG 940
    ||| ::::: ||||| |||||
436 .....LysGluTrpLeuLeuAsp..... 441
941 CTGGAGATACCCATTCAGTATCTACGAACCCAGTCACAAATGGGAAATAC 990
    ||||| ||||| ||||| |||||
442 ..PheAspAsnValSerIleAsnAlaGlnAspProLeuPheAsnAsnThr 457
991 TCTTTTAACGACGAT..AATTAATGCGACAGAAAAATCAATCCCAACA 1037
    ::::: ||||| ::::: |||||
458 GlyIleAsnAspGluPheProThrGlyThrGlyIleAlaSerAspProly 474
1038 TGAACAAATTCCTGCTAATAGATTAAACACGACCGTTCATTTGG 1087
    ::::: ||||| ::::: |||||
474 slyAsnSerGluLeuYThrThrLeuThrAsnThrThrIleSerAsnT 491
1088 TT.....AATGTTTCTTATCCGACAGACAGACA 1116
    ::| ||||| ||||| |||||
491 yrlLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507
1117 GAACCTGTTATCTGCTGCAGGTGGTGTCAACAGTTATGACCCAGACT 1166
    ||||| ||||| ||||| |||||
508 .....ValAsnSer.....SerIle 512
1167 GAATTAATGAGAAATATTTCTTATTTGACGAAAGAAAGCGCAATTTGA 1216
    ||||| ||||| ||||| |||||
512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1217 TACTTACACGACACATCAATCAACAGTCTGAGAGATTATTTCCAAAGA 1266
    ||| ::::: ||||| |||||
527 .....GlnArgGlyGlyValGlnIleAspGly 536
1267 GATTTTACGGCTCGCTGAAATTAACGAAACTTGGCAGCGCGGGCGCT 1316
    ||||| ||||| ||||| |||||
537 AspIleThrSerLysGlyGlyAsnLeuThrIleYThrSerGlyGlyTrpVa 553
1317 TCATATCGTAGAAGACAGTACCTTACTTGAAAGTAAAGCGCGGTGCA 1366
    ||| ::::: ||||| |||||
553 lAspValHisLysAsnIleThrLeu..... 561
1367 ACGACCGCTGTCCAAAATCGCAAAAGCAGCTGACAGTTCAAGCCAAA 1416
561 ..... 561
1417 GGGGAAACCAAGGCTCGATCAGCGTGGCGACGCTACAGTCATTTTGA 1466
    ::||| ||||| ::::: |||||
562 .....AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe.. 575
```

```
1467 TCAGCAGGACGAGCATTAAGCAAAAAACAAGCCTTATGTAATCGGCT 1516
    ::::: ||||| ::::: |||||
576 ..GluGlyGlyAsnAsnLysAlaIleAspAlaIleAlaAsnAlaLysIle... 590
1517 TGTTCACGCGCAGCGGTACGTCGCACTGAAATGCCGTAATATCACTTCAAC 1566
    ::::: ||||| ::::: |||||
591 .....ValAlaGlnGlyThrValThrIleThrGlyGlnGlyLys..... 603
1567 CCCGCAAACTCTTTTCGCTTCCAGCGTATTAACCATGAGTGAAGGCGCATGA 1616
    ||||| ::::: ||||| |||||
604 .....AspPheAlaGlnAsnAsnValSerLeuAsnGln 614
1617 GCATTGCTTTCTTCCACCGTATTAACAAATACCGATGAAGGCGCATGA 1666
    | ||||| ||||| |||||
614 Y.....ThrGlyLysGlyLeuAsnI 621
1667 TTGTC.....AACCAATCAACAGCAAGAAATCCACCGTTAC 1704
    ||| ::::: ||||| |||||
621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
1705 ATTACAGGCAATTAAGATATTTGCTACAAACCGCAATTAACAACAGCTTGA 1754
    ||||| ||||| ||||| |||||
638 lIleSerGlyAsnIleThrIleAsnGlnThr..... 647
1755 TAGCAAAAAAGAAATTTGCTTACAAAGGTTGGTTGGCAGAAAGATACG. 1803
    ::::: ||||| ::::: |||||
648 ..ThrArgLysAsnThrSerTyr.....TrpGlnThrSerHisAspSerH 662
1804 .....ACCAAAACGAACGGGCGGCTGACACTT 1830
662 lStrpAsnValSerAlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPhe 678
1831 GTTTCACGACCCCGCGCAGAACCGCACCTCGTCTT..... 1869
    ::| ::::: ||||| |||||
679 lIleYsTrpIleSerSerAsnSerLysGlyLeuThrThrGlnTrpArgse 695
1870 ..TCGCGCGCAACAATTTAAACGGCACATCAACGCAAAACGCAAC 1918
    ||||| ||||| ||||| |||||
695 rSerAlaGlyValAsnPheAsnGly.....ValAsnGlyAsnM 708
1919 TGTTTTTCAGCGGCGACACCAACCGCACGCTCAATCATTTAAAGCAC 1968
    ::::: ||||| ::::: |||||
708 eSerPhe..... 710
1969 CATTTGTCGCAAAAGAGCGCATTCCTCGCGGCAATCTGTGGAGCAA 2018
    ::| ||||| ||||| |||||
711 .....AsnLeuLysGlnGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGGATCAACCGCAC.....TTTAAAGCGCAA 2050
    |::: ||||| ::::: |||||
725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741
2051 ACTTCCAAATTAAGCGGCGACGCGGTGTT..... 2082
    ||||| ||||| ||||| |||||
741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleYThrAlaAsn 757
2083 ...TCCCGCAATGTTGCCAAAGTGAA...GGCGATTTGCAATTTGAGCAA 2126
    ||||| ::::: ||||| ||||| |||||
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774
2127 TCACGCGCCAGCAGATTTTGGTGTGGCAGCCGATCAAAAGCCACAAACT 2176
    | ||||| ::::: ||||| |||||
774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783
2177 GTACAGCTTCGAGCTGAGCGGTCTGACAAATTTGTGCGAAAAACCAT 2226
    ::||| ||||| ||||| |||||
784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATTAAGTATTTCTTCAATGACTAAGACCGACATCAGCGGCA 2276
    ::| ::::: ||||| ||||| |||||
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815
```

```

2277 TGTGATCTTGGCGATCAGCTCATTTAAATCTACAGGGCTTGGCAGAC 2326
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2327 TCAAGGGCAATCTTAGTGCATATGGCGATACAGCTTATACAGTACGCCAC 2376
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
832 euGIyGLYAsnValThrLeuGIyGLY..... 840
2377 AACGCCACCCAAACGCAACCTTAGCCTCGTGGCAATGGCCCAAC 2426
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
841 .....GlnAsnSerSerSerSerIleThrGlyAsn.....IleTh 852
2427 ATTATATACAGC.....ACATTAACGGCAACATCGGCTTGG 2467
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
852 rIleGIYAsnAlaAlaAsnValThrLeuGIAlaAsnAlaProAsnG 869
2468 GCAATGCTTCAATTATCTAAGCGCACGCGGTACAAACGGCGCTGTG 2517
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
869 LngIn.....AsnIleArgAspArgValIleYsLeuGIYSerLeu 882
2518 ACGCTTTCGGCAACGCTAAGCAACGTAAAGCATTCGCACTCAACGG 2567
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
883 LeuValAsnGIYSerLeuSerIleThrGIYAsnAlaAspIleYsGI 899
2568 TTAATGCTCCCTAGGCCATTAAGCGAGTATTCAT.....TTGAAA 2608
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
899 YAsnLeuThrIleSerGIYSerIleAlaThrPheYsGIYsLeuThrArgAsp 916
2609 GCAGCGGCTTACCGCAAAATCAGCGCGCAAGATACGATACGCAATTAC 2658
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
916 hrLeuAsnIleThrGIYAsnPheThrAsnAsnGIYThrAlaGIAlaAsn 932
2659 TTAAGACAGCGAATGACGCTGCGCTCAGACAGCAAGATTAAGCAATT 2708
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
933 IleThrGIYAsnValValYsLeuGIYAsnValThrAsnAspGIYAspLe 949
2709 AAACCTTACACAGCGCAACCATTAACATTCGCGCTATGCCACGATG 2758
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
949 uAsnIleThr..... 953
2759 CGCGAGGGCGCAACCGCGAGTGCAGACGATGCGCGCGCGCTTCG 2808
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
954 .....HisAla 955
2809 GCGCGTTCGCGCGCTTCCCTATATACGT...TACACCGCAACTTCGG 2854
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
956 LysArgAsnGIYAsnArgSerIleIleGIYAspIleIleAsnYsGI 972
2855 TAGAATCCGCTTTCACACAGCTGACGCTAAGCAAACTTAAGAGGTGAC 2904
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
972 YSerLeuAsnIleThrAspSerAsnAsnAlaGIAlaIleGIY 989
2905 GGAACATTCGCTTATGTGGAACCTTCGCTACGCGACGCAACAATT 2954
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
989 YAsnIleSerGIYAsnGIYAsnLeuThrIleSerSerAspYsI 1005
2955 GAAGCTGGCGAAAGTTCGAAAGCACTTACCTTGGCGTCAACAATA 3004
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1005 eAsnIleThr..... 1008
3005 CCGGCAACGAACTGCAAGCCTCGAACAAATGACGTAGTGAAGAA 3054
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1009 .....LysGIAlaIleThrIleYsGIYIle 1017
3055 GACACAAACCGCTGTTC.....GAAACCTTAATTTCACCT 3092
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1018 AspGIYAsnGIYAsnSerSerSerAspAlaThrSerAsnAlaAsnLeuThr 1034
3093 GCAAAACGAAACGTCGAT.....GCGGCGCGCT 3121
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1034 eYsThrIleYsGIYAsnLeuThrGIYAsnAspLeuSerIleSerGIYphea 1051
3122 GCGGTACCAATCATCGCAAAACGCGC...GAGTTCGCGCTGCATTAAT 3168
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1051 snYsAlaGIAlaIleThrAlaYsAspGIYAsnAspLeuThrIleGIYAsn 1067
3169 CCGGTCAAAAGAAAGAGCTTTCGCAAACTCGCGCAAGCGCAGAAAGCAA 3218
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1068 Ser.....AsnAspGIYAsnSerGIYAlaGIAla.. 1077
3219 AAAACAGCGCGAAACAAACAGCGCGCAAACTTTAGCGCTGATTCGG 3268
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1078 LysThrValThrPheAsnAsnValYs.....AspSerIleSer 1092
3269 CCGGCGCGGATGCGGAAAGACAGAAAGCTTGGCGCAACGCGGCGG 3318
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1092 LAspGIYAsnValThrLeuAsnSerIYsValYsThrSerSer 1108
3319 CAGCGACCGCGGGAATGTGCGCATTAATGACGCGAGAGAGAAAGAAA 3368
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1109 AsnGIYGIYArgGIYSer..... 1114
3369 ACGGTGACGCGGATTAAGACACCGCTTGGCGAAACGCGGAAAGCG 3418
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1115 .....AsnSerAspAsnAspThrGIYLeu..... 1122
3419 AAACCGCGCGGCTACACCGCCTTCGCCGCGCGCGCGCGCGG 3468
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1123 .....ThrIleThrAlaYsAsnValGIYValAsnYs 1133
3469 GATTTCGCGCACTGCAACCCCAACCGCACCCCAACCGCAGCGACT 3518
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1134 AspIleThrSerLeu..... 1138
3519 GATCAGCGCTTATGCAATACGCTTGAATGAATTTCCGCAAGCTCA 3568
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1139 .....LysThrValAsnIleThrAlaSerGIYsValThrThr 1152
3569 ACAGCGTTTTCGCCGTACAGACGAATTAAGCCGCTATTGGCGGAAG 3618
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1153 AlaGIYSerThrIleAsnAlaThrAsnGIYsAlaSerIleThrThr 1169
3619 CGCCGACAGCGCTTGGACAAACGCGCATTCGCGGACCAACACACAC 3668
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1169 sThrGIYAspIleSerGIYThrIleSerGIYAsnThrValSerValSer 1186
3669 TTGCG...AAGATTTCGCGCGCTACCGCCCAACAAACGCACTGCGCA 3715
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1186 LAspValAspLeuThrThrIleYsSerGIYsIleGIYAlaYsSer 1202
3716 TCGGTATGCAAAAACCTCGGACGCGCGCGCTGCGCATTCGTTTCG 3765
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1203 GIYGIYAlaAsnValThrSerAlaThrGIYThrIleGIY.. 1216
3766 CACAACCGGACGCAAAACCTTCGACGAGCGGATGGCA..... 3805
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1217 ThrIleSerGIYAsnThrValAsnValThrAlaAsnAlaGIYAspLeu 1233
3806 .....ACTGCGACGCGCTTGGCCACGCGG 3829
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1233 hrValGIYAsnGIYAlaGIYIleAsnAlaThrGIYAlaIleThrLeu 1249
3830 CGCTTTCGCGCAATACGCAATCGACA.....GGTTACATTCGCAATC 3873
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1250 ThrAlaThrGIYAsnThrLeuThrThrGIYAlaGIYSerSerIleThrSe 1266
3874 AGCGCGCGCGGCTTTAGCAGCGGCACTTTCAGCGGCTGAGGAG 3923
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1266 rThrYsGIYGIYValAspLeuAla.....GlnAsnGIYSerIleAla 1281
3924 CAAATTCGCGCGCGCTGCTCATTAAGCAATTAAGCAGCATACCGG 3973
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1281 LAspGIYSerIleAsnAlaAlaAsnValThrLeuAsnThrThrGIYThrLeu 1297
3974 CCGGTTCGCGGATTCGCAATCGAAACGCAATGCGCGCAACGCGCTAT 4023
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```


474 slySasSerGluLeuLysThrThrLeuThrAsnThrIleSerAsn 491
1088 TT AATGTTCTTATCCGACGACGACA 1116
491 yrlEulysAsnAlaTrpIleMetAsnIleThrAlaSerArgLysLeuThr 507
1117 GAACCTGTTATCATCTCTGAGGTGCTGCAACAGTTATGACCCAGACT 1166
508 ValAsnSer SerIle 512
1167 GAATAATGAGAAATATTTCTTATGACAGGAAAGGCGAATTGA 1216
512 ehsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly 526
1217 TACTTACGACAACATCAATCAAGGTGCTGAGAGATTATATTCCAGCA 1266
527 GlnArgGlyGlyGlyValGlnIleLeuSpgly 536
1267 GATTTACGGTCTCGCTCAAAATAAGCAACTGGCAAGCGCGGCGCT 1316
537 AspIleThrSerLysGlyGlyValAsnLeuThrIleTyrSerGlyGlyTrpVal 553
1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAAGCGCGTGGCA 1366
553 lAspValHisLysAsnIleThrLeu 561
1367 ACGACCGCTGTCCAAATCGGCAAGGACGCTGACGTTCAAGCCAAA 1416
561 561
1417 GGGGAAMCAAGCTCCATCAGCTGGCGGACGATCAGTCAATTTGGA 1466
562 AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe... 575
1467 TCAGCAGCGAGACGATTAAGGCAAAACAGCTTTAGTGAATCGGCT 1516
576 GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle... 590
1517 TGTGTCAGCGGAGGCTACGCTGCAAGTATGCCGATATCAGTTCAAC 1566
591 ValAlaGlnGlyThrValThrIleThrGlyGlyLys... 603
1567 CCCGACAACTCTATTGGGCTTGGCGGCGGAGCTTGGATTAAACGG 1616
604 AspPheArgAlaAsnAsnValSerLeuAsnGly 614
1617 GCATTGCGTTTGTCCACCGATTCAAATACCGATGAGGGGGGAGCA 1666
614 Y ThrGlyLysGlyLeuAsnIle 621
1667 TTGTC AACCAATCAAGACAAAGATCCACCGTTAC 1704
621 IleIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
1705 ATTACGACATTAAGATATGTGTACAAACCGGCATTAACAACAGCTTGA 1754
638 IleSerGlyAsnIleThrIleAsnGlnThr 647
1755 TAGCAAAAAGAAATGGCTACAAACGTTGGTGGCAGAGAAGATAGC 1803
648 ThrArgLysAsnThrSerTyr TrpGlnThrSerHisSpsThr 662
1804 ACCAAACGAAAGCGGCGCTCAACCTT 1830
662 lStrpAsnValSerAlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPhe 678
1831 GTTTACACGCCCGCGAGAGACCGCACCTGCTGCTT 1869
679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSer 695
1870 TCCGGCGGAGCAAAATTAACGGCAACATCCGCAAAACGCGCAAC 1918
695 rSerAlaGlyValAsnPheAsnGly ValAsnGlyLysAsn 708
1919 TGTTTTCAGGCGGACGACCAACCGGACGCTCAATCATTTAAACGAC 1968
708 eSerPhe 710
1969 CATTTGTCGCAAAAAGAGGCAATTCCTCGCGGGGAAATCGTGTGACAA 2018
711 AsnLeuLysGlnGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGCATCAACCGACA TTTAAACGGGAAA 2050
725 ngluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741
2051 ACTTCGCAATTAAGAGCGGACAGCGGTGCTT 2082
741 snIleThrAlaThrGlyGlyGlySerValIlePheAspIleTyrAlaAsn 757
2083 TCCGCAATGTTGCCAAAGTGAAA...GGCATTTGGCATTTGACGAA 2126
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774
2127 TCAGCCGCAACGACTTTTGTGTCGACCGCATCAAGCCACACAATCT 2176
774 nglValAlaAsnPheThrLeu AsnSerHis... 783
2177 GTACACGCTGCGACTGACGCGGTCTGACAAATTTGTCTGCAAAAACAT 2226
784 ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATAAGATGATTCCTCATTTGACTAGACCAACCATCAAGCGCAA 2276
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysSpsPheTyr 815
2277 TGTGATCTTGGCGATCAGCTCATTTAAATCTCACAGGCGTTCACAC 2326
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 832
2327 TCAACGCAATCTTAGTGCAATGGCGATACAGTTATACAGTCAGCGCAC 2376
832 engLysGlyAsnValThrLeuGlyGly 840
2377 AACGCCACCCAAACGGCAACCTTAGCTCTGCGCAATGCCCAACGAC 2426
841 GlnAsnSerSerSerSerIleThrGlyAsn...IleThr 852
2427 ATTAAATCAAGCC ACATTAAAGCCAAACACATCGGCTTCG 2467
852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAlaProAsn 869
2468 GCAATGCTTCATTTAATCTAAGGACGACGCGGTACAAAACGACGACT 2517
869 GlnGln AsnIleArgAspArgValIleLysLeuGlySerLeu 882
2518 AGCTTTCCGCAACGCTAAGGCAACGTAAGCATTCGCACTCAACG 2567
883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGly 899
2568 TAATGTCTCCCTAGCGATTAAGCAGTATTCAT TTGAAA 2608
899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysTrpArgAsp 916
2609 GCAGCGCTTTTACCGGCAAAATCAGCGGCGGCAAGATCCGATTAAC 2658
916 hrlEulysAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932
2659 TTTAAAGACAGCAATGAGACGCTGCGCTCAGCGACGGAATTAAGCAAT 2708
933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949
2709 AAACCTTGACAAAGCCACCATTAACACTCAATTCGCGCTATGCCACGAG 2758
949 uAsnIleThrThr 953

```

2759 CGGCAAGGGCGCAACCGGAGTGCAGATGCGCGCGCGCTGCG 2808
954 .....HisAla 955
2809 CGCGCTTGGCGCTTCCCTATATCCGT...TACACCGCCACCTGCG 2854
956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysL 972
2855 TAGAATCCGCTTCAACACGCTGACGGTAACGGCAATTAACGGCTAG 2904
972 YserLeuAsnIleThrAspSerAsnAspAlaGluIleGlnIleGlyG 989
2905 GGAACATTCGCTTATGTGGAACCTCGGCTACCGACGACAAAT 2954
989 LysnIleSerGlnLysGluGlyAsnLeuThrIle..SerSerAspLysI 1005
2955 GAAGCTGGCGGAAGTTCCGAAGCACTTACACTTGGCGTCAACATA 3004
1005 eAsnIleThr..... 1008
3005 CCGGCAACGACCTGCAAGCCTGCAACATGTGCGTAGTGAAGAGAAA 3054
1009 .....LysGlnIleThrIleLysLysLysIle 1017
3055 GACACAAACCGCTGTC...GAAACCTTAATTTACCTC 3092
1018 AspGlyGlnAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThr 1034
3093 GCAAAACGAACACGTCGAT.....GCCGGCGCT 3121
1034 eLysThrLysGluLeuLysLeuThrGlnAspLeuSerIleSerGlyPheA 1051
3122 GGGGTTCACACTCATCCGCAAGAGCG...GAGTTCCGCGCTCAAT 3168
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067
3169 CCGGTCAAGACAGAGCTTCCGCAAACTCGCAAGGAGGAGAGACCAA 3218
1068 Ser.....AsnAspGlyAsnSerGlyAlaGluAla... 1077
3219 AAAACAGCGCGAAAAGACAAACGCGCAACGCTTGACGCGCTGATTGCG 3268
1078 LysThrValThrPheAsnAsnValLys.....AspSerLysIleSer 1092
3269 CCGGCGCGCATGCGTCAAAAGACAGAAACGCTTGCCGACCGCGCG 3318
1092 LaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSer 1108
3319 CAGCGAGCGGGGAAATGTCGCAATTATGACAGCGGAGAGAGAGAAAA 3368
1109 AsnGlyGlyArgGluSer..... 1114
3369 ACGGGTGGAGCGGATTAAGACACGCGCTTGCGGAACAGCGGAGCGG 3418
1115 .....AsnSerAspAsnAspThrGlyLeu..... 1122
3419 AAACCGCGCGCTACACCGCGCTTCCCGCGCGCGCGCGCGCGG 3468
1123 .....ThrIleThrAlaLysAsnValGluValAsnLys 1133
3469 GATTTCGCGCACTGCAACCCCAACCGGACCGGACCGGCGGACCT 3518
1134 AspIleThrSerLeu..... 1138
3519 GATGCGCGTATGCAATACGCTTGAGTAATTTCCGCCACGCTCA 3568
1139 .....LysThrValAsnIleThrAlaSerGlnLysValThrThr 1152
3569 ACAGCGTTTGGCGCTACAGAGCAATTAAGACCGGATTTGCGGAAGAC 3618
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1169
3619 CGCGCGAACGCGCTTTGGACAAAGCGGCTCGGAGACACCAACACTACG 3668

```

```

1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSer 1186
3669 TTGCG...AGATTTCGCGCGCTACCGCCCAACAAACGACCTCGCCAAA 3715
1186 LaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202
3716 TCGGTATGCAAAACCTCGGACGCGCGCGCTGCGCATCTGTTTCG 3765
1203 GlyLysAlaAsnValThrSerAlaThrGlyThrIleGlyLys..... 1216
3766 CACAACCGGACCGAAACACCTTCGACGACGCGATCGGCA..... 3805
1217 ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1233
3806 .....ACTCGGACGCGCTGCCACGCGG 3829
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaIleThrLeu 1249
3830 CCGTTTTCGCGCAATACGCGATCGACA.....GTTCTACATCGCATC 3873
1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerIleThrse 1266
3874 AGCGCGGCGCGGCTTTAGCAGCGCGACGCTTCAACACGCGATCGAG 3923
1266 rThrLysGlyGlnValAspLeuLeuAla.....GlnAsnGlySerIleA 1281
3924 CAATATCCGCGCGCGCTGCAATACGATTCAGCATTCAGGACGATCCGCG 3973
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297
3974 CCGTTTCGCGGATTCGCGATTCGACCGCACGATCGCGGACGCGCTAT 4023
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuAl 1314
4024 TTGCTGCCAAA 4033
1314 eAsnAlaLys 1317
seq_name: /cgn2_6/plodata/1/laa/5B_COMB.pep:US-08-728-470-4
seq_documentation_block:
: Sequence 4, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEO ID NO: 4
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1477 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-728-470-4

alignment_scores:
 Quality: 282.50 Length: 1388
 Ratio: 0.428 Gaps: 61
 Percent Similarity: 47.550 Percent Identity: 19.380

alignment_block:

US-09-303-518D-649 x US-08-728-470-4

Align seg 1/1 to: US-08-728-470-4 from: 1 to: 1477

```

175 GAAATATAAGCAAGTTGCGAGTCGGGCGAAGATATTGAGTTACAA 224
    ::::::::::::::::::::
219 LysasnGluGlyValIleSerValasnGlySerIleSerIleu.... 233
    ::::::::::::::::::::
225 CAAAAAAGGGAGTTGTGCGCAATCAATGCAAAAGCC..... 264
    ||::::::::::::::::::
234 .....LeuAlaGlyIleThrIleSerAspIleIleA 246
    ::::::::::::::::::::
265 ..CGAGATGATTTTCTGTGTCGCGTAACGCGCGCGCGCTTG 312
    ||| ||| ::::::::::::::::::::
246 snProthIleThrIleSerIleAlaIleProGluasnGluAlaValasn 262
    ::::::::::::::::::::
313 GTGGCGCATCAATATATTGTGAGCGTGCAATACGCGGCTATAACA 362
    ::::::::::::::::::::
263 LeuGlyasp.....IlePheAlaGlyIleGlyAsnIleAs 274
    ::::::::::::::::::::
363 CGTT.....GATTTGTCGCGAAGAA 385
    ||| ::::::::::::::::::::
274 nValArgAlaIleThrIleArgasnGlyIleSerIleAspSerV 291
    ::::::::::::::::::::
386 GAAATCCGATCAATCGTTTACTTATAAATTGAAACGAAATTAAT 435
    ::::::::::::::::::::
291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGly 307
    ::::::::::::::::::::
436 TATAAAGCAGCG.....ACTAAAGCCATCTTATGCGCG 470
    ::| | ::::::::::::::::::::
308 AlagIuIleGlyValIleSerAlaGlnasnGlnIleAlaLysGly 324
    ::| | ::::::::::::::::::::
471 CGATTATCATATGCCGCGTTGCAATAATTGTCACAGATGCAACCTG 520
    | ::| ::::::::::::::::::::
324 LysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341
    ::| ::| ::::::::::::::::::::
521 TTGAATAGCAGATTATATGATGCGGCGAATATATCATCAAAATAT 570
    ::| | ::::::::::::::::::::
341 leaSpLeuSerGlyLysGluGlyIleuThrIleuGlyLysAspGlu 357
    ::| | ::::::::::::::::::::
571 TACCTGACCGTTCGATTATGGGCGAGCGAATATGCGGATTCGA 620
    ||| ||| ::::::::::::::::::::
358 Arg.....GlyGluGlyLysasnGlyIleGlnIleuAl 368
    ::| | ::::::::::::::::::::
621 TGAAGATGAGCCCAATACCGCAAAATTCATATATTCGCAAGT.... 666
    ::| | ::::::::::::::::::::
368 alLysLysThrSerLeuGlyLysSerThrIleAsnValSerGlyLysG 385
    ::| | ::::::::::::::::::::
667 .....GCGTATTCCTGG.....CTCGTTGTGTCGC 690
  
```

```

385 IuLysGlyGlyPheAlaIleValITrGlyAspIleAlaLeuIleAspGly 401
    ||| | ::| | ::| | ::| |
691 AATACCTTTGCAACAAATGATGATGCGTGTGCGACAGTCAACTTAGTAG 740
    ||| | ::| | ::| | ::| |
402 AsnIleasnIleAlaIleGlySerGly..... 409
    ::| | ::| | ::| | ::| |
741 TGAATAATTAACATAGCCCATATGTTTATACCAACAGAGGCTCAT 790
    ::| | ::| | ::| | ::| |
410 .....AspIleAlaLysThrGlyGlyPheV 418
    ::| | ::| | ::| | ::| |
791 TTGGCGAGTGGCTCACCAGTATTTATGATGATCCCAAAAGCAAAAG 840
    ||| | ::| | ::| | ::| |
418 alGluThrSerGlyLysAspLeuPheIleLysasp..... 429
    ::| | ::| | ::| | ::| |
841 TCGTTAATTAATGGGTATTCGAACGCGCAACCCCTATATAGAAAAAG 890
    ||| | ::| | ::| | ::| |
430 .....AsnAlaIleValAspAla..... 435
    ::| | ::| | ::| | ::| |
891 CAATGGCTTCACCTGTTGCTAAGATGTTGTTATGTAATCTTTCG 940
    ||| | ::| | ::| | ::| |
436 .....LysGluThrPheLeuAsp..... 441
    ::| | ::| | ::| | ::| |
941 CTGAGATACCATTCAGTATTCATACGAACACGTCMAAATGGAATAC 990
    ||| | ::| | ::| | ::| |
442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnIleThr 457
    ::| | ::| | ::| | ::| |
991 TCTTTTACGAGCAT...AATAATGCGACAGAAAAATCAATGCCAACA 1037
    ::| | ::| | ::| | ::| |
458 GlyIleAsnAspGluPheProThrGlyThrGlyAlaIleSerAspProly 474
    ::| | ::| | ::| | ::| |
1038 TGAACACAATTCCTGCTAATAGATTAAACACGACGTTCAATGT 1087
    ::| | ::| | ::| | ::| |
474 LysAsnSerGluLeuLysThrIleThrIleThrIleSerAsnT 491
    ::| | ::| | ::| | ::| |
1088 TT.....AATGTTCTTTATCCGACACCAACA 1116
    ::| | ::| | ::| | ::| |
491 LysLeuLysAsnAlaTrpThrMetLeuIleThrAlaSerArgLysLeuThr 507
    ::| | ::| | ::| | ::| |
1117 GAAACGTTTATCATCTCGAGTGTGTCACACAGTTATCCAGCCAGANT 1166
    ||| | ::| | ::| | ::| |
508 .....ValAsnSer.....SerI 512
    ::| | ::| | ::| | ::| |
1167 GAATAATGAGAAATATTTCTTATATGACGAAGAAAGCGAATGA 1216
    ||| | ::| | ::| | ::| |
512 easnIleGlySerAsnSerHisLeuIleuHisSerLysGly..... 526
    ::| | ::| | ::| | ::| |
1217 TACTTACGACAAATCAATCAAGGTGCTGAGAGATTATATTTCCAAAGA 1266
    ||| | ::| | ::| | ::| |
527 .....GlnArgGlyGlyValGlnIleAspGly 536
    ::| | ::| | ::| | ::| |
1267 GATTATAGCGTCTCGCCGTAATAAATGCAAACTTGGCAAGCGGCGGT 1316
    ||| | ::| | ::| | ::| |
537 AspIleThrSerLysGlyLysLeuThrIleThrIleThrIleGlyIleTrp 553
    ::| | ::| | ::| | ::| |
1317 TCATATCACTGAAAGACATACGCTTACTTGAAAGTAAACGCGCGGCAA 1366
    | ::| ::| | ::| | ::| |
553 laeValHisLysAsnIleThrLeu..... 561
    ::| | ::| | ::| | ::| |
1367 ACAGACGCGCTGTCCAAATCGGCAAAAGCACCGCTGACGTTCAACCAA 1416
    ::| | ::| | ::| | ::| |
561 ..... 561
    ::| | ::| | ::| | ::| |
1417 GGGGAAACCAAGGCTCGATCAGCGTGGCGAGCGATGATTTTGA 1466
    ::| | ::| | ::| | ::| |
562 .....AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe.. 575
    ::| | ::| | ::| | ::| |
1467 TCAGCAGCGACAGATAAAGCAAAACCAAGCTTTAGTGAATCGGCT 1516
    ::| | ::| | ::| | ::| |
576 ..GluGlyGlyAsnAsnLysAlaArgAspAlaIleAlaIleAlaLysIle 590
    ::| | ::| | ::| | ::| |
1517 TGGTACGCGGCGAGGCTACGCTGCAACTGAATCCGATATATCATTCAC 1566
    ::| | ::| | ::| | ::| |
  
```

591ValAlaGlnGlyThrValThrIleThrGlyGlyLys..... 603
1567 CCCGACAACTCTATTGGCTTGGCGGGAGCTTGGATTAAACGG 1616
604AspPheArgAlaSerValSerLeuAsnG 614
1617 GCATTGCTTGTTCACCGCTATTCAAAATACCGATGAAGGGCGGTGA 1666
614 Y.....ThcGlyGlyLeuAsnI 621
1667 TTGTC.....AACCAATCAAGACAAAGATCCAGCTTACC 1704
621 IleIleSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
1705 ATTACAGCAATTAAGATATTGTACACCGCAATTAACAACAGCTTGA 1754
638 IleSerGlyAsnIleThrIleAsnGlnThr..... 647
1755 TAGCAAAAAGAAATTTGCTTACAAAGGCTTGGTGGCGAGAAGATACG 1803
648 ThrArgLysAsnThrSerTyr.....TyrGlnThrSerHisAspSerH 662
1804ACCAAAAGCAAGGGCGGCTACACTT 1830
662 IStrPasnValSerAlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPhe 678
1831 GTTTACCAAGCCCGCGCAGAGACCGCACCTGCTGCTT..... 1869
679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695
1870 .TCCGGCGGAAATTTAAAGCGCAATTCACGCAAAACAAAGCGCAAC 1918
695 rSerAlaGlyValAsnPheAsnGly.....ValAsnGlyAsnM 708
1919 TGTTTTACGGCGCAGACCAACCGCACGCTTCAATTTAAAGAC 1968
708 eLSerPhe..... 710
1969 CATTGGTGCAGAAAAGAGCGATTCCTCGCGGGAATTCGTGGAGCAA 2018
711AsnLeuLysGlnGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGCATCAACCGCAC.....TTTAAAGCGCAAA 2050
725 nGlnAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla..A 741
2051 ACTTCCAAATTAAGGCGGACAGCGGT..... 2082
741 snIleThrAlaThrGlyGlySerValPhePheAspIleTyrAlaAsn 757
2083 ..TCCCGCAATGTTGCCAAAGTAA...GGCGATTGGCATTTGACAA 2126
758 HisSerGlyArgGlyAlaGlnLeuLysMetSerGlnIleAsnIleSerAs 774
2127 TCAGCCCAACAGCTTTTGTGTGCGACCGCATCAAGCCACACATCT 2176
774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783
2177 GTACAGCTTCGACTGACGAGGCTGACAAATTTGTGCAAAAACCAT 2226
784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATAAGTGATTTGCTTCAATGACTAAGACCGACATCAGCGCAA 2276
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTyr 815
2277 TGTGATCTTGGCCATCAGCTCATTTAAATTCACAGGCGCTTCCACAC 2326
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 832
2327 TCAAGCGCAATCTTAGTCAAAATGGCGATACAGCTTATACAGTACGAC 2376
832 euGlyGlyAsnValThrLeuGly..... 840

2377 AACGCCACCAAAAGCGCAACCTTAGCCTGTTGGCAATGCCAAGCAAC 2426
841GlnAsnSerSerSerIleThrGlyAsn.....IleTh 852
2427 ATTTAATCAAGCC.....ACATTAAAGCGCAACATCGCTTGG 2467
852 rIleGlnLysAlaAlaAsnValThrLeuGlnLysAlaAsnAlaProAsnG 869
2468 GCAATGCTTCAATTTAATCAAGCAACGCGCTCAAAAGCGCAGTGTG 2517
869 InGln.....AsnIleArgAspArgValIleLysLeuGlySerLeu 882
2518 ACGCTTTCGGGCAACGCTAAGCAAAAGTAAAGCAATTCGACATCAACG 2567
883 LeuValAsnGlySerLeuSerLeuThrGlyGlnAsnAlaAspIleLysG 899
2568 TAATGCTCTCCCTAGCCGATTAAGCAGTATTCAT.....TTGAAA 2608
899 yAsnLeuThrIleSerGlnSerLeuThrPheLysGlyLysThrArgAspT 916
2609 GCAGCGCTTACCGGACAAATCAGCGCGCAAGATACGCAATTACAC 2658
916 hIleuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGlnLeuAsn 932
2659 TTAAGACAGCGAATGAGCGCTGCCGTCAGGCGAGGAATTAAGCAATT 2708
933 IleThrGlnGlyValIleValLysLeuGlyAsnValThrAsnAspGlyAspLe 949
2709 AAACCTTGCAACGCCACCATTAACACTCAATTCGCGCTATGCGCAGATG 2758
949 uAsnIleThrThr..... 953
2759 CGGACAGGGCGCAAAACCGCAGTGGCAGAGATGCCGCGCGCTTGC 2808
954HisAla 955
2809 CGCGGTTCGCGCGCTTCCCTAATTCCT...TACACCGCAATCTTGG 2854
956 LysArgAsnGlnArgSerIleIleGlyLysPheIleAsnLysLysG 972
2855 TAGAATCCGCTTCAACACGCTGACGTAAGCGCAATTTGAACGCTGAG 2904
972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGlnIleGlnIleGlyG 989
2905 GGAACATTCGCTTATGTCGAACTCTTCGCTACCGCAGCAACAAAT 2954
989 LysAsnIleSerGlnLysGlnLysAsnLeuThrIle..SerSerAspLysI 1005
2955 GAAGCTGGCGGAAGTCCGAAGCACTTACACTTGGCGGTCAACATA 3004
1005 eAsnIleThr..... 1008
3005 CCGGCAACGACACTGCAAGCTCGAACAATTCAGCGTAGTGAAGAAA 3054
1009LysGlnIleThrIleLysLysLysIle 1017
3055 GACACAAACCGCTGTC.....GAAACCTTAATTTACCT 3092
1018 AspGlyGlnAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrI 1034
3093 GCAAAAGCAACACGTCGAT.....GCCGGCGCGT 3121
1034 eLysThrLysGlnLeuLysLeuThrGlnAspLeuSerIleSerGlyPheA 1051
3122 GCGCTTACCAACTCATCCGCAAGACGCG...GAGTTCCGCTGCATAT 3168
1051 snLysAlaGlnIleThrAlaLysAspGlyArgAspLeuThrIleGlyLysn 1067
3169 CCGGTCAAGAACAGAGCTTTCGCAAACTCGCAGACGCAAGAACCA 3218
1068 Ser.....AsnAspGlyAsnSerGlyAlaGlnAla... 1077


```

3219 AAACAGCGCGGAAAAAGACAAACGCGCAAGCGCTTGACGCGCTGATTGGG 3268
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1078 LysThrValThrPheAsnValLys.....AspSerLysIleSerA 1092
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3269 CCGGCGCGGATGCCCTGCAAAAAGACAAACGCTTGCGGACCGCGCGG 3318
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1092 LaaSpGlyHisAsnValThrLeuAsnSerLysValLysThrSerSer 1108
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3319 CAGCAGCGCGGGAATGTCGGCATTTGACAGCGGAGAGAGAAAAA 3368
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1109 AsnGlyGlyArgGluSer..... 1114
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3369 ACGGCTGACAGCGGATTAAGACACCGCTTGCGAAMACGCGGCAAGCGG 3418
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1115 .....AsnSerAspAsnAspThrGlyLeu..... 1122
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3419 AAACCGCGCGGCTAACACCGCGCTTCCCGCGCGCGCGCGCGCGG 3468
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 .....ThrIleThrAlaLysAsnValGluValAsnLys 1133
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3469 GATTTCGCGCACTGCACACCCCAACCGACCGCCACCGCGGACCT 3518
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1134 AspIleThrSerLeu..... 1138
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3519 GATCAGCGCTTATGCGCAATAGCGGTTGATGAAATTTCCGCGCACGCTCA 3568
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1139 .....LysThrValAsnIleThrAlaSerGlyLysValThrThr 1152
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3569 ACAGCGTTTCGCGCTACAGAGCAATTAAGCGCGGATTTGCGCGAGAC 3618
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrLys 1169
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3619 CGCGCAACGCGCTTGACAAAGCGGATCCGCGGACCAACCAACTACG 3668
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3669 TTCCG...AAATTTCCCGCGCTACCGCGCAACCAACGACCTGCGCAAA 3715
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 IatThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3716 TCGGTATGCAAAAAACCTCGCGAGCGCGCGCGGATCGGATCTTTTCG 3765
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGly..... 1216
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3766 CACACCGGACGCAAAACACCTTGACAGCGGATCGGCA..... 3805
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1217 ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1233
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3806 .....ACTGGCGACGCGCTTGCGCGCGGCG 3829
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3830 CCGTTTCGCGCATACGCGCATCGACA.....GGTCTACATCGCGCATC 3873
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1250 ThrIatThrGlyAsnThrLeuThrThrGlyAlaGlySerIleThrSe 1266
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3874 AGCGCGCGCGCGGCTTTTACAGCGGCGAGCTTTACAGCGGATCGGAG 3923
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1266 rThrLysGlyGlyValAspLeuAla.....GlnAsnGlySerIleA 1281
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3924 CAATAATCGCGCGCGCTGCTGCTGATTCAGCGCATTCAGCGCATACCGG 3973
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1281 IatGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3974 CCGGTTTCGCGCGGATTCGCGCATCAACCGCATCGCGCAACGCGCAT 4023
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1314
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4024 TTCGTCGCAAA 4033
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1314 eAsnAlaLys 1317
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/1/laa/5B_COMB.pep:US-08-617-697-4

```

seq_documentation_block:
; Sequence 4, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Matlare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-4

alignment_scores:
      Quality: 282.50      Length: 1388
      Ratio: 0.428      Gaps: 61
      Percent Similarity: 47.550      Percent Identity: 19.380

alignment_block:
US-09-303-518D-649 x US-08-617-697-4 ..
Align seg 1/1 to: US-08-617-697-4 from: 1 to: 1477

175 GAAATAAAGGCAAGTTTCAGTCCGCGGCAAGATATTGACGTTTACAA 224
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 LysAsnGluGlyValIleSerValAsnGlySerIleSerLeu..... 233
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 CAATAAAGGAGAGTTGGTGGCAATCAATGCAAAAGGC..... 264
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 .....LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 ..CCGATGATTGATTTCTGTGTGTCGCGGTACGCGGCGCGCATG 312
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 snpThrThrIleThrLysSerIleAlaIleAlaProGluAsnGluAlaValAsn 262
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 GTGGCGATCATATATTGTGAGCGTGGCACATACGCGCGCTATAACA 362
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 LeuGlyAsp.....IlePheAlaLysGlyGlyAsnIleAs 274
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

725 nglusnlethsrSerLysProleuProleuPheleuAla...A 741
726 ACTTCCAAATTAAGCGGAGCGGCGT... 2082
741 snllethralaThrcllygllySerValPhepheaSpillethryAlaasn 757
2083 ...TCCGCCAATGTTGCCAAAGTAAA...GGCGATTGGCATTTGGACAA 2126
758 HlsSerGlyArgGlyAlaGluLeuLysMetSerGluLleasnLleSerAs 774
2127 TCACGCCCAAGCAGTTTGGTGGTCGACCGCATCAAGCCACACCAATCT 2176
774 nglYAlaasnPhethrLeu...AsnSerHis... 783
2177 GTACACGTTGCGACTGACGGGTTGACAAATTTGTCGAAAAACCAAT 2226
784 ...ValArgGlyAspAspAlaPheLysIleasn...LysAspLeuThrIle 798
2227 ACCGACGATTAAGTATTGCTTCATTTGATGAAGCCGACATCAGCGCAA 2276
799 AsnAlaThrAsnSerAsnPheserLeuArgLlnThrLysAspAspPheTy 815
2277 TGTGATCTTGGCCGATCAGCTCATTTAAATCTCACAGGGCTGCCACAC 2326
815 rAspGlyTyAlaArgAsnAlaIleasnSerThrTyAlaSnLleSerIleL 832
2327 TCAACGGCAATCTTATGTCGAATGAGCGATACAGCTTATACAGTCAGCAC 2376
832 euglGlyAsnValThrLeuGly... 840
2377 AACGCCACCCAAAGCGACCTTACCTCGTGGCAATGCCCAAGCAAC 2426
841 ...GlnAsnSerSerSerSerIleThrGlyAsn...IleTh 852
2427 ATTATATCAGCC...ACATTAACGGGACACATCGGCTTCGG 2467
852 rIleGluLysAlaIleasnValThrLeuGluAlaAsnAlaProasnG 869
2468 GCAATGCTTCATTTATCTAAGCGACACGCGGTACAAACGCGAGTCTG 2517
869 lngln...AsnIleArgAspArgValIleLysleuGlySerLeu 882
2518 ACGCTTTCGGCAACGCTAAGGCAACGATTCGCCACTCAACGG 2567
883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaSpIleLysG 899
2568 TAATGTCTCCTTACCGATAGCGAGTATTCAT...TTGAAA 2608
899 YasnleuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspTr 916
2609 GCAGCCCGCTTACCGCAAAATCAGCGCGGCAAGATAGCGCATTCAC 2658
916 hrLeuAsnIleThrGlyAsnPherAsnAsnGlyThrAlaGluLleasn 932
2659 TTAAGAACAGCGAATGAGAGCTGCCGTCAGCGACGCAATTTAGCAATTT 2708
933 lIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949
2709 AAACCTTGACAAAGCGCACCATTTACACTCAATTCGCTATCGCCAGATG 2758
949 uAsnIleThr... 953
2759 CGGCAAGGCGCAACCGGACGTGCGACAGATGCGCGCGCGCTTCG 2808
954 ...HisAla 955
2809 CGCGCTTCGCGCGCTTCCTATTATTCGT...TACACGCCCAACTTCGG 2854
956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleasnLysGly 972
2855 TAGAATCCCGTTTCMACCGCTGACGCTAAACGGCAATTTGAACGCTCAG 2904
972 YSerLeuAsnIleThrAspSerAsnAsnAlaGluIleGlnIleGlyG 989
2905 GGAACATTCGCCCTTTATGTGGAACCTTCGTACCGCAGCGCAAAAT 2954
989 LysnLleSerGlnLysGluGlyAsnLeuThrIle...SerSerAspLysI 1005
2955 GAAGCTGGCGCAAAAGTCCGAGAGCACTTACCTTGGCGGTCAACAAATA 3004
1005 eAsnIleThr... 1008
3005 CGCGCAACGAACCTGCAAGCCTGCAACAAATTTGAGTAGAGAGAA 3054
1009 ...LysGlnIleThrIleLysGlyLe 1017
3055 GACACAAACCGCTGTC...GAAACCTTATTTACCT 3092
1018 AspGlyGluAspSerSerSerAlaThrSerAsnAlaAsnLeuThrI 1034
3093 GCAAAACGACACGTCGAT...GCCGGCGCT 3121
1034 elYsThrLysGluLeuLysLeuThrLysPleuSerIleSerGlyPhea 1051
3122 GCGTTACCACTCATTCGCCAAAGACGGC...GAGTTCCGCTGCATTAAT 3168
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067
3169 CCGGTCAAGAACAGACGCTTCCGACAAACTGCGACAGCGACAGCCAA 3218
1068 Ser...AsnAspGlyAsnSerGlyAlaGluAla... 1077
3219 AAAACAGCGGCAAAAGACACCGCGCAAGCCTTGAGCGGTGATGCGG 3268
1078 ...LysThrValThrPheAsnAsnValLys...AspSerLysIleSerA 1092
3269 CCGGCGCGCATGCTCGCAAAAGACAGAAACGCTTCCGACACCGCGCG 3318
1092 LaAspGlyLysAsnValThrLeuAsnSerLysValLysThrSerSer 1108
3319 CAGCGAGCGGGGGAATAATGCGCATTTATGAGCGCGAGAGAGAGAAAA 3368
1109 AsnGlyGlyArgGluSer... 1114
3369 ACGGTGCGCGGATTAAGACACCGCTTGGCGAAACAGCGCGAAGCGG 3418
1115 ...AsnSerAspAsnAspThrGlyLeu... 1122
3419 AAACCGCGCGGTACACCGCTTCCCGCGCGCGCGCGCGCGCGCG 3468
1123 ...ThrIleThrAlaLysAsnValGluValAsnLys 1133
3469 GATTGCGCGCACTGCAACCCCAACCGCAGCCCAACCGCAGCGGACCT 3518
1134 AspIleThrSerLeu... 1138
3519 GATCAGCGCTTATGCAATAGCGGTTGAGTGAATTTTCCGACAGCTCA 3568
1139 ...LysThrValAsnIleThrAlaSerGluLysValThrThr... 1152
3569 ACAGCGCTTTCGCCGTACAGAGCAATTTAGACCGGTATTTCCGAAAGC 3618
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrly 1169
3619 CGCCGCAACGCGGTTTGGACAAAGCGGATCGCGACACCAACACTACCG 3668
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186
3669 TTGCG...AAGATTTCCGCGCTTACCGCCAAACAGACCTGCGCCAAA 3715
1186 lathrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202
3716 TCGGATGAGAGAAAAACCTGCGAGCGCGCTGGGACATCCTGTTTCG 3765
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly... 1216

```

```

3766 CACACCGGACCCAAACACCTTCGACGAGCGATCGCA..... 3805
1217 .ThrIleSerGlyAsnThrValAsnValThrIleAsnAlaGlyAspLeuT 1233
3806 .....ACTGCGACGCGCTTGCCACGCGG 3829
1233 hTValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaIleThrLeu 1249
3830 CCGTTTGGGCAATACGCGATCGACA.....GGTTTCATACGCGCATC 3873
1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266
3874 AGCGCGGCGCGGGGTTTACGACGCGCCTTCAGACGCGCATCGAGG 3923
1266 rThrLysGlyGlnValAspLeuLeuAla.....GlnAsnGlySerIleA 1281
3924 CAAATCCGCGCGCGCTGCTGCATTACGGCATTCAGGACGATACCGCG 3973
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297
3974 CCGGTTTCGGCGGATTCGGATCGAACCGACATCGCGCGCATCGGTAT 4023
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValII 1314
4024 TTCGTCCAA 4033
1314 eAsnAlaLys 1317

```

seq_name: /cgn2.6/ptodata/1/1aa/6B.COMB.pep:US-08-719-641-4

seq_documentation_block:

Sequence 4, Application US/08719641
Patent No. 6218141

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-625

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 4:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-4

```

```

alignment_scores:
  Quality: 282.50      Length: 1388
  Ratio: 0.428        Gaps: 61
  Percent Similarity: 47.350      Percent Identity: 19.380

```

alignment_block:

US-09-303-518d-649 x US-08-719-641-4 ..

Align seg 1/1 to: US-08-719-641-4 from: 1 to: 1477

```

175 GAAATTAAGGCAAGTTTCGTGATGTCGTAACGGCGTGGCGCATTG 224
225 CAAATAAGGCGAGTTGTCGCAATCAATGCAATAAGCC..... 264
234 .....LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246
265 ..CCGATGATGATTTTCTGTGTGTCGTAACGGCGTGGCGCATTG 312
246 snPrThrIleThrIleThrSerIleAlaAlaProGluAsnGluAlaValAsn 262
313 GTGGCGCATCAATATATTGTGAGCGTGGCACATACGGCGCTTAACAA 362
263 LeuGlyAsp.....IlePheAlaLysGlyGlyAsnIleAs 274
363 CGTT.....GATTGTGTCGCGAAGGAA 385
274 nValArgAlaIleThrIleArgAsnGlnGlyLysLeuSerAlaAspSerV 291
386 GAAATCCCGATCAACATCGTTTACTTATAATTGTGAACGGAATAT 435
291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGlnLysIu 307
436 TATAAAGCAGG.....ACTAAGCGCATCTTATGGCGG 470
308 AlaGluIleGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyI 324
471 CGATTATCATATGCGCGCTTGCATTAATTTGTCACAGATCGACAACCTG 520
324 LysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341
521 TTGAATGACAGTATATGATGGCGGAATATATATGATCAATAATAT 570
341 leAspLeuSerGlyLysGlnGlyGlyGluThrLysLeuLysGlyAsp 357
571 TACCTGACCGTGTGCTATGCGAGCGACGACGATATGCGCATCGA 620
358 Arg.....GlyGlnGlyLysAsnGlyIleGlnLeuAl 368
621 TGAAGATGAGCCCAATACCGGAAAGTTCATATATGCAAGT.... 666
368 alLysLysThrSerLeuGlnLysGlySerThrIleAsnValSerGlyLysG 385
667 .....GCGTATCTTGG.....CTGCTTGCTGCG 690
385 LysGlyGlyPheAlaIleValIleThrGlyAspIleAlaIleAspLys 401
691 AATACCTTGCAAAATGATGATGAGTGTGGCAGCACTTATGAGTAG 740
402 AsnIleAsnAlaGln...GlySerGly..... 409
741 TGAATAATTAACATAGCCCATATGTTTATACCAACGAGAGCGCTAT 790
410 .....AspIleAlaLysThrGlnGlyGlyPheV 418

```

```

791 TTGGGACAGTGTCCACCAATGTTATCTATGATGCCCAAAAGCAAAAG 840
      |||||
418 aLgLUthSerGlyHisAspLeuPheIleLysASP..... 429
841 TGGTTAATTAATGGGTAATTGCAAAAGGCGCAACCCCTATATAGAAAAAG 890
      |||||
430 .....AsnAlaIleValAspLa..... 435
891 CAATGGCTCCAGCTGCTGCTTAAGATGTGTCTATGATGAATCTTGTG 940
      |||||
436 .....LysGluTrpLeuLysP..... 441
941 CTGGAGATACCCATTCAGTATTTCTAGCAACACGTCACAAATGGGAATAC 990
      |||||
442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457
991 TCTTTTAACGACGAT..AATATGGCACAGAAAAATCAATGCCAACA 1037
      |||||
458 GlyIleAsnAspGluPheProThrGlyThrGlyAlaIleSerAspProLy 474
1038 TGAACACATTCCTGTGCTTAATAGATTAATAACAGACCGTCAATGT 1087
      |||||
474 sLysAsnSerGluLeuLysThrLeuThrAsnThrThrIleSerAsnT 491
1088 TT.....AATGTTCTTTATCCGACAGACAGACA 1116
      |||||
491 yLLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507
1117 GAACTGTTTATCATGCTGCACAGTGTCTCAACAGTTTTCACCCCACT 1166
      |||||
508 .....ValAsnSer.....SerIle 512
1167 GAATATGAGAAAAATATTCTTATATGACAGAAAAAGCGCAATTGA 1216
      |||||
512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1217 TACTTACACGACACATCATCAGTGTCTGAGAGATTATATTTCACAGA 1266
      |||||
527 .....GlnArgGlyGlyValGlnIleAspLy 536
1267 GATTTTACGTCGTGCGCTGAAAAATACAGAACTTGGCAAGCGCGGCGT 1316
      |||||
537 AspIleThrSerLysGlyLysAsnLeuThrIleThrSerGlyGlyTrpVa 553
1317 TCATATCATGTAAGACACTACCTTACTTGGAAAGTAACGCGCTGCAG 1366
      |||||
553 LAspValHisLysAsnIleThrLeu..... 561
1367 ACGACCGCTGTCCAAAAATCGCAAAAGCAGCTGCACGTTCAAGCCAAA 1416
      |||||
561 ..... 561
1417 GGGAAAAACAAGGCTGATCAGCGTGGCGAGGCTACAGTCATTTTGA 1466
      |||||
562 .....AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe. 575
1467 TCACAGGACAGACGATAAGGCAAAAAACAAGCCCTTATAGTAATCCGCT 1516
      |||||
576 .GluGlyLysAsnLysAlaArgAspAlaIleAsnAlaLysIle..... 590
1517 TGGTCAGCGGAGGAGTACGTCAGTGAATGCCGATATCACTTCAAC 1566
      |||||
591 .....ValAlaGlnGlyThrValThrIleThrGlyGlyLys..... 603
1567 CCCGACAAACTCTATTTCGCTTCGCGGGGAGCTTGGATTAAACGG 1616
      |||||
604 .....AspPheArgAlaAsnAsnValSerLeuAsnGly 614
1617 GCATTCGCTTCTGTTCCACCGTATTCAAAATACCGATGAAGGCGGATGA 1666
      |||||
614 y.....ThrGlyLysGlyLeuAsnIle 621
1667 TTGTC.....AACCAATATCAAGCAAAAGAAATCCACCGTTAC 1704
      |||||
621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
1705 ATTACAGCAATAAAGATATTGCTACAAACCGGCAATTAACAACAGCTTGA 1754
      |||||
638 IleSerLysAsnIleThrIleAsnGlnThr..... 647
1755 TAGCAAAAAAGAAATTGCTTACAAACGGTTGGTTGGCGAGAAAGATACG 1803
      |||||
648 ..ThrArgLysAsnThrSerTyr.....TrpGlnThrSerHisAspSerH 662
1804 .....ACCAAAACGAACCGCGCGCTCAACCTT 1830
662 lStrAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678
1831 GTTTACAGCCCGCGAGAGACCGACCCCTGCTCTT..... 1869
679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695
1870 .TCGCGCGGACAAATTAAACGGCAACATCAGCAAAACAGCGCAAC 1918
      |||||
695 rSerAlaGlyValAsnPheAsnGly.....ValAsnGlyAsnM 708
1919 TGTTTTCACGCGGACAGACCAACACCGACGCTTACATCATTTAAACGAC 1968
      |||||
708 eLsePhe..... 710
1969 CATTTGTCGCAAAAGAGGCAATTCCTCGCGGGAAATCGTGTGGACA 2018
      |||||
711 .....AsnLeuLysGlnGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGATCAACCGACA.....TTTAAAGCGGAAA 2050
      |||||
725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla..A 741
741 snIleThrAlaThrGlyLysSerValPhePheAspIleTyrAlaAsn 757
2083 ...TCCGCGCATGTTCGCAAAAGTAAA...GGCGATTGGCATTTGAGCA 2126
      |||||
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774
2127 TCACGCCCAACAGAGTTTGGTGTCCACCGCATCAACGCCACACAATCT 2176
      |||||
774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783
2177 GTACAGCTTGGAGCTGAGCGGTCTGACAAATTTGTTCGAAAAAACCAT 2226
      |||||
784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATTAAGTGAATTGCTTATTGACTTAAGACCGCATCAGCGGCA 2276
      |||||
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815
2277 TGTGATCTTGGCATCAGCTCATTTAATATCTCAACAGGCGTGGCACAC 2326
      |||||
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832
2327 TCACAGGCAATCTTACTGCAAAATGGCGATACAGTTATACAGTACGCCAC 2376
      |||||
832 euGlyLysAsnValThrLeuGlyGly..... 840
2377 AACGCCAACCAAAACGCAACCTTAGCTTCGTGGGCAATGCCCAAGCAAC 2426
      |||||
841 .....GlnAsnSerSerSerIleThrGlyAsn.....IleTh 852
2427 ATTTAATCAAGC.....ACATTAAAGGCAACACATCGGCTGG 2467
      |||||
852 rIleGlyLysAlaAlaAsnValThrLeuGluAlaAsnAlaProAsnG 869
2468 GCATGCTTCATTTAATCTTAAGCAGACACCGCTCAAAAGCGCAGTGTG 2517

```

```

869  IngIn.....AsnIleArgAspArgValIleLysLeuGlySerLeu 882
      ::::: |||:::|||||
2518  ACCTTCCGCGCAACGCTAAGCAACGCTAAGCATTCGCCACTCAACGG 2567
      ::::: |||:::|||||
883  LeuValAsnGlySerLeuSerLeuThrGlyLysAsnAlaAspIleLysGly 899
      ::::: |||:::|||||
2568  TAATGTCCTCCATGCCATAGGAGCATATTCAT.....TTTGAA 2608
      |||:::|||||
899  yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 916
      |||:::|||||
2609  GCAGCCGCTTACGCGCAAAATACGCGCGCAAGATACGCGATTAAC 2658
      ::::: |||:::|||||
916  hIleuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932
      ::::: |||:::|||||
2659  TTAAAAACAGCAGCAATGAGCGCTGCCGTCAGCAGCAATATGCAAT 2708
      ::::: |||:::|||||
933  IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspLysPle 949
      ::::: |||:::|||||
2709  AAACCTTGACAAAGCCACCATTAACATCAATTCGCTATCGCCAGATG 2758
      |||:::|||||
949  uAsnIleThr..... 953
2759  CGCGAGGCGCGCAAAACGCGAGTGCAGATGCGCGCGCGCTTCG 2808
      ::::: |||:::|||||
954  .....HisAla 955
2809  CGCGCTTCGCGCCCTTCCTATATCCGT...TACACGCCCAACTCGG 2854
      ::::: |||:::|||||
956  LysArgAsnGlnArgSerIleIleGlyLysPleIleAsnLysLysGly 972
      ::::: |||:::|||||
2855  TAGATCCCGTTTCACACGCTGACGCTAAACGCAATTAAGACGTCAG 2904
      ::::: |||:::|||||
972  ySerLeuAsnIleThrAspSerAsnAsnAlaGluIleGlnIleGly 989
      ::::: |||:::|||||
2905  GGAACATTCGCTTATGTGCGAACTCTGGCTACCGCAGCGCAAAAT 2954
      ::::: |||:::|||||
989  LysAsnIleSerGlnLysGlnLysLeuThrIle..SerSerAspLysI 1005
      ::::: |||:::|||||
2955  GAAGCTGGCGGAAAGTCCGAAAGCATTAACCTTGGCGTCAACATA 3004
      ::::: |||:::|||||
1005  eAsnIleThr..... 1008
3005  CGCGCAACGAACTGCAGACCTCGAACATTAAGTGAAGCAAAA 3054
      ::::: |||:::|||||
1009  .....LysGlnIleThrIleLysLysGlyIle 1017
      ::::: |||:::|||||
3055  GACAAACAACCGCTGTC.....GAAACCTTAATTTCACCC 3092
      ::::: |||:::|||||
1018  AspGlyLysAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrI 1034
      ::::: |||:::|||||
3093  GCAAAAGCAACAGTCAT.....GCCGCGCGT 3121
      ::::: |||:::|||||
1034  eLysThrIleGlnLysLeuLysLeuThrGlnAspLeuSerIleSerLys 1051
      ::::: |||:::|||||
3122  GCGCTTACCACTCATCCGCAAAAGCGC..GAGTTCGCGCTGCATAAT 3168
      ::::: |||:::|||||
1051  snLysAlaGluIleThrAlaLysAspLysArgAspLeuThrIleGlyAsn 1067
      ::::: |||:::|||||
3169  CGGCTCAAGAAACAAGACTTTCGCAAAACTCGCAAGGCAAGCAAC 3218
      ::::: |||:::|||||
1068  Ser.....AsnAspLysAsnSerGlyAlaGluAla.. 1077
      ::::: |||:::|||||
3219  AAAACAGCGGAAAAAGCAAAAGCGCAAAAGCTTGACGCGCTGATTCG 3268
      ::::: |||:::|||||
1078  LysThrValIleThrPheAsnAsnValLys.....AspSerLysIleSe 1092
      ::::: |||:::|||||
3269  CGGCGCGGAGTCCGCGCAAAAGCAAAAGCTTGCGCAAGCGCGCGG 3318
      ::::: |||:::|||||
1092  LaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSer 1108
      ::::: |||:::|||||
3319  CAGGAGGCGGGGAAATGTCGGCATTTATGAGGCGGAGGAAGAAAA 3368
      ::::: |||:::|||||

```

```

1109  AsnGlyIleArgGluSer..... 1114
3369  ACGGTCGAGCGGCGATTAAGACACCGCTTGCGCAAAACGCGCAAGCG 3418
      ::::: |||:::|||||
1115  .....AsnSerAspAsnAspThrGlyLeu..... 1122
      ::::: |||:::|||||
3419  AAACCGGCGCGCTACACCGCGCTTCCCGCGCGCGCGCGCGCGG 3468
      ::::: |||:::|||||
1123  .....ThrIleThrAlaLysAsnValGluValAsnLys 1133
      ::::: |||:::|||||
3469  GATTTCGCGCAACTGCACACCCCAACCGGACGCCCAACGCGAGACT 3518
      |||:::|||||
1134  AspIleThrSerLeu..... 1138
3519  GATCAGCGCTTATGCCAATAGCGGTTTGATGAATTTCCGCGACGCTCA 3568
      ::::: |||:::|||||
1139  .....LysThrValAsnIleThrAlaSerGluLysValIleThr..Thr 1152
      ::::: |||:::|||||
3569  ACAGCGTTTTCGCGCTACAGACGAAATTAACCGCTATTTGCGCAAG 3618
      ::::: |||:::|||||
1153  AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrLys 1169
      ::::: |||:::|||||
3619  CCGCCGACGCGCGCTTGACAAAGCGGCAATCCGCGGACACCAACACTAC 3668
      ::::: |||:::|||||
1169  sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValIleValIle 1186
      ::::: |||:::|||||
3669  TTCGC...AGAATTCCGCGCGCTACCGCAACCAACGCGCGCGCAAA 3715
      ::::: |||:::|||||
1186  laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSe 1202
      ::::: |||:::|||||
3716  TCGGTATGCAAAAAACCTCGCGAGCGGCGCGCTGCGCATCTGTTTCG 3765
      ::::: |||:::|||||
1203  GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216
      ::::: |||:::|||||
3766  CACAAACGCGGCAAAACACCTTCGACGCGCGCATCGCA..... 3805
      |||:::|||||
1217  ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1233
      ::::: |||:::|||||
3806  .....ACTCGCACGCGCTTGCACGCGG 3829
      ::::: |||:::|||||
1233  hrValGlyAsnGlyAlaGluIleAsnAlaThrGluIleAlaIleThrLeu 1249
      ::::: |||:::|||||
3830  CCGTTTTCGCGCAATTCGCAATCGCA.....GGTTCTACATCGGCATC 3873
      ::::: |||:::|||||
1250  ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266
      ::::: |||:::|||||
3874  AGCGCGGCGCGGTTTACAGCGGCAAGCTTTCAGACGCGCATCGAGG 3923
      ::::: |||:::|||||
1266  rThrLysGlyGlnValAspLeuLeuAla.....GlnAsnGlySerIle 1281
      ::::: |||:::|||||
3924  CAAATTCGCGCGCGCGCTGCTGCATTAAGCGCATTAAGCAGCATACGG 3973
      ::::: |||:::|||||
1281  laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297
      ::::: |||:::|||||
3974  CCGGTTTCGCGGATTCGCGATTCGGAACGCGCATCGCGCAAGCGCTAT 4023
      ::::: |||:::|||||
1298  ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValI 1314
      ::::: |||:::|||||
4024  TTCGTCGCAAA 4033
      ::::: |||:::|||||
1314  eAsnAlaLys 1317
      ::::: |||:::|||||

```

seq_name: /cgn2_6/ptodata/1/lae/5A_COMB.pep.us-08-038-682-2

seq_documentation_block:

; Sequence 2, Application US/08038682

; Patent No. 5549897

; GENERAL INFORMATION:

; APPLICANT: BARENKAMP, STEPHEN J

; APPLICANT: ST. GENE III, JOSEPH W

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

; OF NON-TYPEABLE HAEMOPHILUS

; NUMBER OF SEQUENCES: 8

727 nlystf.....SerSerLeuAsnTyAlaSer. 736
1199 AAGGAAAGGGAATGTACTACTACCAACATCATCAAGGCGTGA 1248
737PheAsnGlyAsnIleSerValSerGlyGly 746
1249 GGATTATATTTCCAAAGGAGATTTTACGTCGCCCTGAAATAACGAAC 1298
747 Gly.....SerValAspPheThrLeuLeuAlaSerSerSerAnva 760
1299 TTGGCAAGCGCGCGCTT.....CATTCAGTG 1327
760 L...GlnThrProGlyValIleAsnSerLysTyPheAsnValSer 776
1328 AAGACAGTACCGTCTTGTGAATAAAGCGC..... 1359
776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792
1360 ...GTGGCAAGCAGC...CGCCTGCCAAATCGCAAGCAGCAGTGA 1403
793 SerIleGlyLysAspLeuThrLeuAsnAlaThrGlyLysAsnIleThrLe 809
1404 CGTTCAACCCAAAGCGCAAAAC.....C 1426
809 ulenGlnValGlnGlyThrAspGlyMetIleGlyLysGlyIleValAla 836
1427 AAGCGTCAGTACGCGTGGCGCAGCTACAGTATTGGATTCACAGCAGCA 1476
826 yslYAsnIleThrPheGlnGlyLysAsnIleThrPhe..... 838
1477 GACGATAAAGGCAAAACAAACGCTTAGTGAATCGGCTTGTCACGG 1526
839GlySerArgLysAlaValThrGluLe.....GluGln 849
1527 CAGGGTACGCTGCAACGATGCGGATTAATCACTTCACCCGACAAAC 1576
849 yAsnValThrIleAsnAsnAsnAlaAsn..... 856
1577 TCTATTGCGCTTCGCGCGCAGCTTGGATTAAACGGGCAATCGCT 1626
859ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCACCGTATTCAAAATACCGATGAAGGGCGATGATTCACACCA 1676
867 AspnAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA.....GACAAAGAATCCACCGTACCATTCAGACGAATAAG 1720
883 rGlyAsnLeuThrAlaGlyLysAlaIleValAsnIleAlaGlyAsnLeu 900
1721 ATATTGCTCAACCGGCAAT..... 1740
900 hrValGlnSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
1741AACACAGCTTGATAGCAAAAAAGAAATGGCTTA 1775
917 ValGlnGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLys 933
1776 CAACGCT...TGGTTTGGCAGAAAGATACGACAAACGAAAGCGCGCG 1822
933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....AsnL 947
1823 TCACCTTGTTTACCAGCGCGCGCAGAAAGACCGCACCTGCTGCTTCC 1872
947 euserIleThrThrAsnSerSerThrTyArgThrIleIleSerGly 963
1873 GCGGGAATAATTTAAAGGACATCAACGCAACAAACGCAACGCTT 1922
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTTCAGCGCAGACCAACCGCAGCGCTTACATTTAAACGACCAT 1972
977 ...GluGlySerAspThrGlu.....MetGlnIleGlyLysAsp 989

1973 GGTCCAAAAGAGGGCATTCCTCGCGGGAATCGTGGACACAGCAG 2022
989 aISerGlnLysGlu.....GlyAsnLeuThrIleSerSerAsp 1001
2023 TGGATTCACCGCAGCATTTAAAGCGGAAACCTTCAATTTAAAGCGGACA 2072
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGly.. 1014
2073 GCGGCGTGTTCGCCCAATGTTGCCAACTGAAAGCGCATTCGATTGA 2122
1015ValAspGlyGlnLysAsnSerAspSerAlaLys 1025
2123 GCATTCACCGCAGCATTTTGGTGTGCGCACCGCATCAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACAGTTCGAGTCTGACGGGCTCTGACAAATGTGTGCAAAAC 2222
1032 IleLysThrLys.....GluLeuLys 1038
2223 CATTCACCGCAGTAAAGTATGCTTCATTCATTAAGCCGACATCAGCG 2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGlnIleThrA 1055
2273 GCATGTGATCTTGCAGTACGCTCATTTAAATCTCACAGGCGTTCGC 2322
1055 Lals...AspGlySerAsp.....Leu 1061
2323 ACACGTACACGCAATCTTACTCAATGGCATACACGTATACGTAG 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAlaAlaLysLysValTh 1078
2373 CCACAGCGCCACCAAAAGCAGACCTTAGCTCGTGGCAATGCCAAG 2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..... 1091
2423 CAACATTTAATCAAGCCACATTTAAAGCGCACATCGGCTTCGGCAAT 2472
1092HisLysValThrLeuHisSerLysValGlnThrSerGlySer 1105
2473 GCTTCATTTAATCTAAGCCACGACGCGCTAACAAGCGGAGCTGACGT 2522
2523 TTCCGCGCAAGCGCTAAGGCAACGATTAAGCCATTCGACATCAACGTAATG 2572
1106 AsnAsn...AsnThrGlnAspSerSerAspAsnAlaGlyLeuThrIle 1121
1121 e.....AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132
2573 TCTCCCTACCGCATTAAGCAGTATTCATTTGAAAGCAGCGCTTACG 2622
1132 LeThr...SerHisLysAlaVal.....SerIleSerAlaThrSer 1144
2623 GGCACAAATCAGCGCGCGCAAGGATACGGCATACATTAA.....GA 2666
1145 GlyLeuIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161
2667 CACGCAATGACGCTCCCTCAAGCAGCAGCAATTAAGCAATTTAAACCTTG 2716
1161 nValGlnIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGlnLys 1178
2717 ACACGCGCACCATTCATTCATTCATTCGCTATCGCAGCATCGCGCAGG 2766
1178 eSerGlySerValThrLeuThrAlaThrGlnGlyAlaLeuAlaValSer 1194
2767 GCGCAACCGCGCAGTGCAGACATGCGCGCGCGCGCTGCGCGCTTG 2816
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211
2817 GCGCGCTTCCTATTAATCGGTACACCGCAACTTCGATGAAATCGCTT 2866
1211 rThrLeuAlaGlySerThrIleLysGlyThrGlnSerValThrThrSers 1228


```

2867 TCACACGCTGACGGTAACGGCAATTTGAACGCTCAGGACA..... 2910
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1228 eGInsercIyaspIleGIythrIleSerGIy...GIythrValGIu 1243
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2911 TTCGGCTTATGTCGGAACCTCTCGGCTACCGCAAGCAACAAATTTGAAGCT 2960
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1244 ValIySaIaThrGIuSerIeuThrGIuSerIaSerIyIleYs.. 1259
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2961 GCGGAAAGTTCCGAAGGACCTTACACCTTGCGGCTC..... 2997
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1260 .AlaThrThrGIyGIuAlaIaSnValThrSerIaThrGIyThrIleGIyG 1276
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2998 .....ACAAATACCGGCAAGCAAGCTCGAAGCTCGCAACATTTG 3036
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1276 LyThrIleSerGIyasnThrValaSnValThrAlaIaSnAlaGIyaspIeu 1292
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3037 ACGGTA.....GTGGAAGAAAGACACAA 3062
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1293 ThrValGIyasnGIyAlaGIuIleasnAlaThrGIuGIyAlaIaIaThrIle 1309
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3063 ACCGCTGTCGCAAACTTAATTTACCCCTGCAAAAGCAACGCGATG 3112
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1309 uThrThrSerSerGIyIySerIeuThrThrGIuAlaSerIeHisIleThrs 1326
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3113 CCGGCGCGTGGCGTTACCACTCATCGCAAGAGCGGAGTTCCGCGCTG 3162
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1326 eTAlaIySGIyGIuValaIaSnIeuSerIaGIuIaSpIySerValaIaGI 1342
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3163 CATATCGGTCGCAAAAGCAAGAGCTTCCGCAACATCGGCAAGGCGCA 3212
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1342 ySerIleasnAlaIaSnValThrIleasnThrThrGIyThrIeuThrT 1359
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3213 AGCCAAAACAGAGCGGAAAGCAACGCGCAAAAGCTTG 3253
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1359 hrValIySGIySerIaSnIleasnIaThrSerGIyThrIeu 1372
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: /cgn2_6/prodata/1/1aa/5A_COMB.pep:US-08-302-832-2
seq_documentation_block:
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-2

alignment_scores:
      Quality: 272.50      Length: 1182
      Ratio: 0.452
Percent Similarity: 51.015      Percent Identity: 20.897

alignment_block:
US-09-303-518D-649 x US-08-302-832-2 ..

Align seq 1/1 to: US-08-302-832-2 from: 1 to: 1536

127  GGACACACTTATTTGCGCATCATCACTACATATCATCGGACTTTCGCA 176
      ||:::|||||:::|||||
349  GIGluThrIyIreugIy.....GIyAs 356
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
177  AAATTAAGCAAGTTTGCACTCGGCGCAAGATATTGAGTTTACACA 226
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
356  pGIuArgGIyGIu.....GIyIyasnGIyIleGIuIleValaIaIy 370
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
227  AAAAAGGCGGTTGTGCGCAATCAATGACAAAGCCCGATGATTGAT 276
      ||  :::::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::
370  ys.....ThrSerIeuGIuIySerIeThrIleasn 380
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
277  TTTTCTGTGCTGCGGTACGCGCGGCGCATTTGTCGCGCATATA 326
      ||  :::::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::
381  ValSerGIySGIuIySGIyArgAlaIleValIreGIyaspIleAl 397
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
327  TATTGTG.....ACGTCGCACTA 346
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
397  aleuIleaspGIyasnIleasnAlaGIuIySerGIyaspIleAlaIy 414
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
347  ACGGCGGC..... 354
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
414  hrGIyGIyIreValGIuThrSerGIyHisaspIeuIleIyIyasp 430
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
355  .....TATACAGCTTGATT 371
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
431  AlaIleValaspAlaIySGIuThrIleuaspIeaspIaValSerI 447
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
372  TGTGCGGGA.....GGAAGAAATCCGATCAACATCGTTTACTAT 414
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
447  eAsnAlaGIuThrAlaGIyArgSerasnThrSerGIuaspIyIyT 464
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
415  .....AAATGTGAAGCAAGATATTATATAAGCA 444
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
464  hrGIySerGIyasnSerIaSerThrProIyAsArgasnIySGIuIyS 480
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
445  GGGACTTAAGGCCATCTTATGCGGCGGATTCATATATGATGATG 494
      ||  :::::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::
481  ThrIeuThrAsnThrThrIeuGIuSerIeIleuIySGIuIyIre 497
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
495  TAAATTGTCACAGATCGAGAACCTTTGAATGACCACTTATGATG 544
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
497  alasnIleThrIaasnGIuArg.....IleIyVal 507
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
545  GCGGGAATATATCATCAAAATTAATTAACCTGACCGTTCGTATGCG 594
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
508  AsnSerSerIleasnIeuSerasnGIySerIeuIreIreIreIre 524
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
595  GCAGGAGGC.....AATATTGCGCATC 617
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

524 YargserglylyglyValglulleasnAsnAspIleThrThrlyAspa 541
618 TGATGAAGATGAGCCCATACC.....GGCAAGTT 649
541 spThrArgglyAlaAsnLeuThrIleTyrSerglylyTyrPalaAspVal 557
650 CATATCATATTCGAAGCTGCTATTCCTGCTGCTGCTGCAATACCTTT 699
558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573
700 GCACAAAT.....GGATCA.....GCTGG 719
574 LysGlnAspIleAlaPheGlyLysGlySerAsnGlnValIleThrGlygl 590
720 TGGCAGCTCAACTAGTAGTGAATAAATAACATAGCCCATATAGCTT 769
590 nGlyThrIleThrSerGlyAsnGlnLys.....GlyP 601
770 TTTTACCAAGAGAGGCTCATTTGGCGACAGTGCATCCCATGTTATTC 819
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617
820 TATATGCTCCCAAAAGCAAAAGTGTATTAATGCGGTATTCGAACGGG 869
618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGlnGlyTh 634
870 CAACCCCTATATAGCAAAAGCAAT...GGCTTCACCTGCTTCTGTAAG 916
634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651
917 ATTGTTCTATGATGAATCTTGGCTGGAGATACCAT.....TCAGTA 960
651 luserGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666
961 TTTACCAACACCTCAAAATGGGAATCTCTTTTAACACATATATA 1010
667 SerLeuAsnValSerGlnSerGlyLupheAsnLeuThrIleAspSerAr 683
1011 TGGCACA.....GGAATAATCATGCCAATAACATGAACACATCTCTC 1051
683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700
1052 TGCCCT...AATGATTAATAAACACGAACCGTTCATATTGTTATGTTCT 1098
700 lSerPheAsnLysAspThrThrPheAsnValGlnArg...AsnAlaArg 715
1099 TTAATCCGAGACAGCAAGAACCTGTTATCATCTGCTGAGGTGTCGCA 1148
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAs 727
1149 CAGTTATCGACCCAGACTGATTAATGAGAAATATTTCTTTATTCAGC 1198
727 nLysTyr.....SerSerLeuAsnTyrAlaSer. 736
1199 AAGGAAAGGCGAATTGATCTTACCGACCAATCAATCAAGGTCGCGA 1248
737PheAsnGlyAsnIleSerValSerGlyGly 746
1249 GGATATATTTTCCAAAGAGATTTTACGCTCGCTGGAATAATACGAAC 1298
747 Gly.....SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760
1299 TTGGCAAGCGCGGCGCTT.....CATATCAAGT 1327
760 L...GlnThrProGlyValIleAsnSerLysTyrPheAsnValSerT 776
1328 AAGCAGTACCGTTACTTGAAGATAACGCG..... 1359
776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792
1360 ...GTGGCAAGACG...CGCTGTCCCAAAATCGCAAAAGCAGCTGCA 1403
793 SerIleGlyLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 GGTTCACCCAAAGGCGAAAC.....C 1426
809 uLeuGlnValGlyLysThrAspGlyMetIleGlyLysGlyLeValAlaL 826
1427 AAGCTGCATACCGTGGCGCACGCTACATCATTTTGCATCAGCAGCA 1476
826 ySLysAsnIleThrPheGlnGlyLysIleThrPhe..... 838
1477 GACGATTAAGGCAAAACAGCCCTTTAGTGAATCGCTGTGTCAGCG 1526
839GlySerArgLysAlaValThrGlnIle.....GlnG 849
1527 CAGGGGTACCGTGCACATGAATGCCGATATCATGTTACCCCGACAAAC 1576
849 yAsnValThrIleAsnAlaAsnAlaAsn..... 858
1577 TCTATTTGGCTTTGGCGCGGACCTTGTGATTAACGGCATTTGCTT 1626
859ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCCACCGCTATTCAAAATACCGATGAAGGGCGATGATTCACACA 1676
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA.....GACAAAGAATCCACCGCTTACCATTTACAGCAATTAAG 1720
883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlnLysLeuT 900
1721 AATATGCTACACCGGCAAT..... 1740
900 hrValGlnSerAsnAlaAsnPheLysAlaIleThrAspPheThrPheAsn 916
1741AACACAGCTTGGATGACCAAAAGAAATTCCTTA 1775
917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933
1776 CAACGCT...TGGTTGGCGGAAAGATACGCCAAACGAAAGCGCGCG 1822
933 scLysGlyAlaArgPheLysAspIleAspAsnSerLys.....AsnL 947
1823 TCAACCTGTTTACACGCCCGCGGAGAAAGCAGCACCTGCTGCTTCC 1872
947 euserIleThrThrAsnSerSerThrTyrArgThrIleIleSerGly 963
1873 GCGCGAACAATTTAAACGCGCAACATCACGCAAAACGCGCAACTGT 1922
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTACCGCGCAGACCAACCGCGCAGCCCTACAAATCATTTAAAGCANT 1972
977 ...GlnGlySerAspThrGlu.....MetGlnIleGlyGlyAspV 989
1973 GGTGCAAAAGGAGGCAATTCCTCGGGGGAATGCTGGGACACAGC 2022
989 aISerGlnLysGlu.....GlyAsnLeuThrIleSerSerAsp 1001
2023 TGGATCAACCGCACATTTAAAGCGGAAACTTCAATTAAGCGGCA 2072
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGly.. 1014
2073 GCGCGTGTTCGCCGAATGTTGCCAAGTGAAGGCAATGGCATTTGA 2122
1015ValAspGlyGlnLysSerSerSerAspAlaT 1025
2123 GCAATACCGCCAGACAGTTTTGGTGTGCGACCCCATCAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACACGTTGCGACTGACGCGGTCTGACAAATGTGTGCAAAAAC 2222
1032 IleLysThrLys.....GluLeuLys 1038

```

1309 urthrhSerSergjlylsleuthrhtgrhualaserSerhislethrS 1326
      |||..... ||| .....|||... :
1313 CCGGCGCGGTTCCTTCCACCTCATCGCAAGCGCGAGTCCGCTG 3162
      :: |||::||| ::||| |||||::: ::||| |||
1326 erAlbysllylnValasnlseuSerAlaglnspSlserValaGl 1342
      |||::: ||| :::: |||::: |||:::
1342 ySerlIeasnlAlasvnlhrleuasnThrtGjlythrleuthr 1359
      |||::: ||| :::: |||::: |||:::
1359 hrVallysglySerasnIleasnlAlathrSergjlythrleu 1372
      |||::: ||| :::: |||::: |||:::

seq_name: /cgn2_6/plodata/1/1aa/5B_COMB.pap:US-08-530-198-2

seq_documentation_block:
; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEIN(S)
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-2

alignment_scores:
      Quality: 272.50      Length: 1182
      Ratio: 0.452      Gaps: 60
      Percent Similarity: 51.015      Percent Identity: 20.897

alignment_block:
US-09-303-518D-649 x US-08-530-198-2 ..
Align seg 1/1 to: US-08-530-198-2 from: 1 to: 1536

127 GGACACACTATTTCGGCATCAACTACCAATACTATCCGACGCTTGCCGA 176
|||||.....|
349 GtlygturhtThyTleudly.....Gtlyas 356
|||||.....|

```



```

ADDRESSER: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

alignment_scores:
Quality: 272.50 Length: 1182
Ratio: 0.452 Gaps: 60
Percent Similarity: 51.015 Percent Identity: 20.897

alignment_block:
US-09-303-518D-649 x US-08-469-880-2 ..

Align seg 1/1 to: US-08-469-880-2 from: 1 to: 1536

127 GGACACTTATTTGGCATCACTACCAATATCATCGCGACTTGGCGA 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 G|G|G|U|T|T|Y|L|E|U|G|Y|.....G|L|Y|A|S| 356
177 AAATAAAGCAAGTTTGCAGTCGGGGCGAAGATATTGAGTTTACACA 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 PGLVHRTG|L|Y|G|L|U|.....G|L|Y|S|A|S|N|G|Y|L|E|G|L|N|E|U|A|L|A|Y|S|L 370
227 AAAAAAGGAGCTGTGTCGCAATCAATGACAAAAGCCCGCATATTGAT 276
|||:|||||:|||||:|||||:|||||:|||||:|||||:
370 YS.....Th|S|e|r|L|e|u|G|L|U|Y|S|G|L|Y|S|e|r|T|h|L|e|a|S|N 380
277 TTTTCTGTGTGTCGCTAAGCGCGTGGCGCATTTGGTGGCGCATACATA 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:
381 V|A|S|e|r|G|L|Y|S|G|L|U|Y|G|L|Y|A|R|G|A|L|A|L|E|V|A|L|T|P|G|L|Y|A|S|P|L|I|E|A|I 397
327 TATTGTG.....A|G|C|G|T|G|C|A|C|A|T|A 346
|||||:|||||:|||||:|||||:|||||:|||||:
397 A|L|e|u|I|e|A|S|P|L|Y|S|e|r|L|e|u|A|S|N|A|L|A|G|L|Y|S|e|r|G|L|Y|A|S|P|L|I|E|A|L|A|Y|S|T 414
347 ACGGCGGC..... 354

```

```

414 hnglyGlyPheValG|U|ThrSerGlyH|S|AspLeuPhe|L|e|Y|AspAsn 430
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 .....T|A|T|A|C|A|A|C|G|T|G|A|T|T| 371
431 A|L|A|L|E|V|A|L|A|S|P|A|L|A|Y|S|G|L|U|T|P|L|E|U|S|A|S|P|H|E|S|A|S|N|V|A|S|E|r|I| 447
372 TGGTSCGGA.....G|G|A|G|A|A|A|T|C|C|G|A|T|C|A|C|A|T|G|T|T|T|A|C|T|T|A|T 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 e|A|S|N|A|L|U|T|H|A|G|L|Y|A|R|G|S|e|r|A|S|N|T|H|S|e|r|G|L|A|S|P|L|U|Y|T 464
415 .....A|A|A|T|G|T|G|A|A|C|G|A|T|A|T|T|A|A|G|C|A 444
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 h|N|G|L|Y|S|e|r|G|L|Y|A|S|N|S|e|r|A|L|A|S|e|r|T|H|P|R|O|L|Y|S|A|R|G|A|S|L|Y|S|G|L|U|S|T|H|R 480
445 G|G|G|A|C|T|A|A|G|C|C|A|C|C|T|T|A|T|G|G|G|G|G|G|A|T|C|A|T|A|T|G|C|G|C|G|T|T|G|C|A 494
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 .Th|L|e|u|T|H|A|S|N|T|H|T|H|L|E|U|S|e|r|I|L|E|U|L|Y|S|G|L|Y|H|P|H|E|V 497
495 T|A|A|T|T|G|T|C|A|C|A|G|T|G|C|A|A|C|C|T|G|T|G|A|A|A|T|G|C|A|G|T|T|A|T|G|A|T|G|C|A 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 A|L|A|S|N|I|L|E|T|H|A|L|A|S|N|G|L|N|A|R|G|.....I|L|E|T|Y|V|A|L 507
545 G|G|C|G|A|A|A|T|A|T|G|C|A|A|A|T|A|T|A|T|A|C|C|T|G|A|C|C|G|T|T|C|G|A|T|T|G|G 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 A|S|N|S|e|r|I|L|E|A|S|N|L|E|U|S|e|r|A|S|N|G|L|Y|S|e|r|L|E|U|T|H|L|P|S|e|r|G|L|U|I 524
595 G|C|A|G|C|A|G|C|.....A|A|T|T|G|G|C|G|A|T|C 617
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 Y|A|R|G|S|e|r|G|L|Y|G|L|Y|V|A|L|G|L|U|L|E|A|S|N|A|S|P|L|E|H|T|H|R|G|L|Y|S|A|P|A 541
618 T|G|A|T|G|A|G|A|T|G|A|G|C|C|A|A|T|A|A|C|.....G|C|G|A|A|G|T| 649
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 S|P|T|H|A|R|G|L|Y|A|L|A|S|N|L|E|U|T|H|I|L|E|Y|S|e|r|G|L|Y|T|T|P|V|A|L|S|A|P|A|L 557
650 C|A|T|A|C|A|T|A|T|G|C|A|G|T|G|G|T|T|C|T|G|C|G|T|G|T|G|G|C|A|A|T|A|C|T|T|T 699
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 H|I|S|L|Y|S|A|S|N|I|L|E|S|e|r|L|E|U|L|A|G|L|N|G|L|Y|A|S|N|I|L|E..A|S|N|I|L|E|T|H|A|L 573
700 G|C|A|C|A|A|A|T|.....G|G|A|T|C|A|.....G|G|T|G| 719
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
574 L|Y|S|G|L|A|S|P|L|E|A|L|A|P|H|E|G|L|U|L|Y|S|G|L|Y|S|e|r|A|S|N|G|L|N|A|L|I|E|T|H|G|L|Y|G|I 590
720 T|G|C|A|C|A|G|T|C|A|C|T|A|G|T|A|G|T|G|A|A|A|A|A|T|T|A|A|C|A|T|G|C|C|A|T|T|G|G|T 769
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 N|E|L|Y|T|H|I|L|E|T|H|S|e|r|G|L|Y|A|S|N|G|L|Y|S|.....G|L|Y|P 601
770 T|T|T|A|C|A|C|A|G|A|G|G|C|T|A|T|T|G|G|C|G|A|C|A|G|T|G|C|C|A|C|A|T|G|T|T|A|T|C 819
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 h|E|A|R|P|H|E|A|S|N|A|S|N|V|A|L|S|e|r|L|E|U|A|N|G|L|Y|T|H|G|L|Y|S|e|r|G|L|Y|L|E|U|G|L|N|P|H|E 617
820 T|A|T|G|A|T|G|C|C|A|A|A|G|C|A|A|A|G|T|T|A|T|T|A|T|G|G|G|G|A|T|T|G|C|A|A|C|G|G 869
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
618 T|H|T|H|L|Y|A|R|G|T|H|A|S|N|Y|S|T|Y|A|L|A|L|E|r|H|A|S|N|L|Y|S|P|H|E|G|L|Y|T|H 634
870 C|A|A|C|C|C|T|A|T|A|G|A|A|A|A|A|G|C|A|A|T|.....G|G|T|T|C|A|G|G|T|G|T|C|T|A|A|G 916
634 R|L|E|U|A|S|N|I|L|E|S|e|r|G|L|Y|V|A|L|S|N|I|L|S|e|r|V|E|T|A|L|L|E|U|P|R|O|L|Y|S|A|N|G 951
917 A|T|T|G|T|T|A|T|G|A|T|G|A|A|T|C|T|T|T|G|G|A|T|A|C|C|A|T|.....T|C|A|G|T|A 960
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 I|U|S|e|r|G|L|Y|Y|A|S|P|L|Y|S|.....P|H|E|L|Y|S|G|L|Y|A|R|G|T|H|T|Y|T|T|P|A|S|N|L|E|U|T|H|R 666
961 T|T|C|T|A|C|A|C|A|C|G|T|C|A|A|A|A|T|G|G|A|A|A|T|C|T|T|T|A|C|G|A|C|G|A|T|A|A|T|A 1010
667 S|e|r|L|E|U|A|S|N|V|A|S|e|r|G|L|U|P|H|E|A|S|N|L|E|U|T|H|I|L|E|A|S|S|e|r|A|r 683
1011 T|G|C|G|C|A|A|.....G|G|A|A|A|T|C|A|A|T|C|C|A|A|C|A|T|G|A|C|A|A|T|T|G|C 1051
683 G|G|L|Y|S|e|r|A|S|P|S|e|r|A|L|G|L|Y|T|H|L|E|U|T|H|G|L|N|P|R|O|T|Y|A|S|N|L|E|U|A|S|N|G|L|Y|I 700
1052 T|G|C|C|T|...A|T|A|G|A|T|T|A|A|A|C|A|G|A|A|C|G|T|C|A|A|T|T|G|T|T|A|T|G|T|T|C|T 1098
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

700 leSerPheasnLysaspThrThrPheasnValGluArg... AsnAlaIaArg 715
1099 TTATCCGAGACAGCAAGACACCTGTTATATCATGCTGCAGCGTGTCA 1148
      ::::: |||:::
716 ValAsnPheAspIleIleValaIaProIle..... GlyIleAs 727
1149 CAGTTATGCAACCCAGACATGAAATATGAAAAATATTCCTTATGTCAG 1198
      ||::|||
727 nLysTyr..... SerSerLeuAsnTyrAlaSer. 736
1199 AAGGAAAAGCGCAATTGATCTACTTACCACCAATCAATCAAGCTCTGA 1248
      ::::: |||:::
737 ..... PheasnGlyAsnIleSerValSerGly 746
1249 GGATTATATTTCCAGAGATTTTACGCTCTCGCTGAAATTAACGAAC 1298
      ||| ::::: |||:::
747 Gly..... SerValAspPheThrIleuLeuAlaSerSerSerAsnVal 760
1299 TTGGCAGCGCGGCGCTT..... CATATCAGTG 1327
      ::|||
760 L... GluThrProGlyValIleAsnSerIleTyrPheasnValSerT 776
1328 AAGCAGTACCGCTTACTTGGAAATGAAGGC..... 1359
      |||:::
776 hArgIleSerIleuArgPheIleuArgSerGlySerThrLysThrGlyPhe 792
1360 ...GTGGCAAGAC... CGCCTGTCCAAATCGGCAAGCGACGCTCA 1403
      ::::: |||:::
793 SerIleGlyLysAspLeuThrIleuAsnAlaThrGlyLysAsnIleThrLe 809
1404 CGTTAACCCAAAGGGGAAAC..... C 1426
      ::|||:::
809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826
1427 AAGGCTCATGACGCTGGCGACGCTACGATTTTGATTTGATCAGACAGCA 1476
      ::|||:::
826 yLysAsnIleThrPheGluGlyLysAsnIleThrPhe..... 838
1477 GACGATAAGGCAAAAAACAACCTTATGTAATCGGCTTGTCAGCG 1526
      |||:::
839 ..... GlySerArgLysAlaValThrGluIle..... GluG 849
1527 CAGGGGTACGGTCACTGAATGCCGATATCACTTCAACCCGACAAC 1576
      ||::|||:::
849 yAsnValThrIleAsnAsnAlaAsn..... 858
1577 TCTATTTCGGCTTCGCGCGACGTTTGATTTAAAGCGCATTCGTT 1626
      ::|||
859 ..... ValThrIleuIleGlySerAspPhe 866
1627 TCGTTCCACCGTATTCAAATACGAGTGAAGGGCGATGATGTCAACA 1676
      ::|||:::
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA..... GACAAAGAATCCACCGTACCATTAACGCAATAAG 1720
      :::::
883 rGlyAsnLeuThrAlaGlyLysAsnIleValAsnIleAlaIleLysLeuT 900
1721 ATATTGCTACACCGGCAT..... 1740
      ::|||:::
900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
1741 ..... AACACAGCTTGATATGCAAAAAAGAAATTCGCTA 1775
      ::|||:::
917 ValGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaIal 933
1776 CAACGCT... TGCTTGGCGAGAAAGATAGCAACCAAAACGACGGCGC 1822
      ::|||
933 sGlyGlyAlaIaArgPheLysAspIleAspAsnSerLys..... AsnL 947
1823 TCAACCTGTTTACCAAGCCCGCAGCAAGAAGACCGACCTGCTGCTTCC 1872
      |||:::
947 eSerIleThrThrAsnSerSerThrTyrArgThrIleIleSerGly 963
1873 GCGGACAATTTAAACGCGACATCAGCAAAACAACGCAACTGTT 1922
      ::|||
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTCAGCGGAGACACACCGCGCGCTACATTTAAACGACCATT 1972
      ::|||
977 ...GluGlySerAspThrGlu..... MetGlnIleGlyLysAsp 989
1973 GGTGCAAAAAGAGGCGATTCCTCGCGGGAATCGTGGGCAACAGAC 2022
      |||:::
989 alSerGlnLysGlu..... GlyAsnLeuThrIleSerSerAsp 1001
2023 TGATATCACCGGCACATTTAAAGCGGAAACTTCCAAATTAAGCGGACA 2072
      |||:::
1002 LysIleAsnIleThr..... LysGlnIleThrIleLysAlaGly 1014
2073 GCGGTGTGTTCCCGCAATGTGCAAGTGAAGCGATTTGCGATTGCA 2122
      |||
1015 ..... ValAspGlyGluAsnSerAspSerAspAlaT 1025
2123 GCAATACGCGCCCAAGCAGTTTGTGTGCGACCGCATCAAGCCACACA 2172
      ::|||
1025 hrAsnAsnAlaAsnLeu..... Thr 1031
2173 ATCTGTACAGCTTCGACGACGGGTGACAAATTTGTGCAAAAAAC 2222
      |||
1032 IleLysThrLys..... GluLeu 1038
2223 CATTAACGAGATTAAGTATGCTTCATTTAGTAAGCGACGACATCAGCG 2272
      ::|||
1038 sleuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055
2273 GCATGTGCAATCTTCGCGATCAGCTCATTTAATCTCACAGGCTTGCC 2322
      ::|||
1055 IaLys... AspGlySerAsp..... Leu 1061
2323 ACATCAACGCGCAATCTTATGCAAAATGGGATACAGCTTATACAGTCA 2372
      |||
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysValThr 1078
2373 CCAACAGCCACCCAAAGCGCAACCTTACCTCTGGGCAATGCCCAAG 2422
      ::|||
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..... 1091
2423 CAACATTTATCAAGCACATTTAAAGCGACACATCGGCTTCGGGCAT 2472
      ::|||
1092 ..... HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105
2473 GCTTCATTTAATCTAAGCGACGACGCGGTACAAAGCGAGCTGACGCT 2522
      ::|||
1106 AsnAsn... AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrI 1121
2523 TTCCGCGACAGCTTAAGCAACGTAAGCCATTCGACATCAACGTAATG 2572
      ::|||
1121 e..... AspAlaLys... AsnValThr..... ValAsnAsnAsnI 1132
2573 TCTCCCTAAGCGGATFAGGAGATTTCCATTTGAAAGCGCGCTTTAC 2622
      ::|||
1132 leThr... SerHisLysAlaVal..... SerIleSerAlaThrSer 1144
2623 GGACAAATGACGCGCGGCAAGATACGGCATTAACCTTAAAA..... GA 2666
      |||
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161
2667 CAGCGATGAGCGCTCGCTCAGCGACAGCAATTTAGCAATTTAAACCTTG 2716
      ::|||
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSer 1178
2717 ACAACGCCACATTAACATCAATTCGCTATCGGCGACAGATCGGCGAGG 2766
      ::|||
1178 eSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

```


[illegible]

```

1015 .....ValaspGlygluaAsnSerAspSerAspAlat 1025
1213 GCATATCAGCCCAAGCAGTTTGGTGTGGCACCAGCATCAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACAGTTCGGACGTGGACGGGTCTGCAAAATGTGTGCAAAAAC 2222
1032 ILleYsthrIys.....GlueuY 1038
2223 CATTACCGAGATAAAGATGCTTCATTAGTACAGCCAGCATGACG 2272
1038 sleuThrGluaSpleuAsnIleSerGlyPheAsnIysAlaIleuIleTha 1055
2273 GCATATCGCATCTTGGCATCAGGCTCATTAATCTCACAGGCGTTGCC 2222
1055 Ialys...AspGlySerAsp.....Leu 1061
2323 ACATCAACGCGCATCTTACGCAATGCGCATACAGCTTATACGTAC 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaIysIysValTh 1078
2373 CCACACGCGCACCCAAAAGCGCAACCTTACGCTGGGCGCATGCCAAG 2422
1078 rPheAsnGlnValIysAspSerIysIleSerAlaAspGly..... 1091
2423 CAACATTATCAAGCCATTAAGCGCAACACATCGCTGGGCGCAT 2472
1092 .....HisIysValThrLeuHisSerIysValGlnThrSerGlySer 1105
2473 GCCTCATTTATCTACGACGACGCGCGTACAAAGCGGAGTGCAGCT 2522
1106 AsnAsn...AsnThrGlnAspSerSerAspAsnAsnAlaIleuThrIle 1121
2523 TTCGGGACGCTAAGGCAACGTAACCCATTCGCGACTCAAGGTAAG 2572
1121 e.....AspAlaIys...AsnValThr.....ValAsnAsnAsnI 1132
2573 TCCTCCGACGATAGGACGATATTCATTGTTGAAGCCCGCTTAC 2622
1132 IeThr...SerHisIysAlaVal.....SerIleSerAlaThrSer 1144
2623 GGACAAATCAGCGCGGCGCAAGGATACGCACTTAAATAA.....GA 2666
1145 GlyIleuIleThrThrIysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161
2667 CAGGATGAGCGCTGCGCTACGACGAGGAATTAAGCAATTAACCTTG 2716
1161 nValGlnIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178
2717 ACAACGCGCATATACATCAATCCGCTATGCGCCACGATGCGCGAGG 2766
1178 erSerGlySerValThrLeuThrAlaThrGlnGlyAlaLeuAlaValSer 1194
2767 GCGCAAAACGCGAGTGCACAGATGCGCGCGCGCGCTTGCAGCGCTG 2816
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211
2817 GCGCGCTTCCTATTATCCGTTACCGCAACTCGGTGAGTACCCGCT 2866
1211 rThrLeuAlaIysSerThrIleIysGlyThrGluSerValThrThrSerS 1228
2867 TCACACGCGCTGACGTTAAACGCAAAATGAACGCTGAGGACACA..... 2910
1228 erGlnSerGlyAspIleGlyIleThrIleSerGly...GlyThrValGlu 1243
2911 TTCGCGTTATGTCGGAACCTTCGCTACCGCAGCAGCAAAATGAAGCT 2960
1244 ValIysAlaThrGlnSerLeuThrThrGlnSerAsnSerIysIleYs.. 1259
2961 GCGCAAAAGTTCGCAAGGCACTTACACTTGGCGGTC..... 2997

```

```

1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGly 1276
2998 .....AACATACGCGCAACGACCTGCAGAGCTGCAGCAATG 3036
1276 lYThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspleu 1292
3037 ACGGTA.....GTGAGAGGAAAGACACAA 3062
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGlnGlyAlaIleThrle 1309
3063 ACCGCTGCCAAACCTTATTCACCTTCGCAAAAGCAACAGCTGCATG 3112
1309 uThrThrSerSerGlyIysIleuThrThrGluAlaSerSerHisIleThrS 1326
3113 CCGGCGCGGCGGTACCACTCATCCGCAAGAGCGGAGTTCGCGCTG 3162
1326 erAlaIysGlyGlnValAsnIleuSerAlaGlnAspGlySerValAlaG 1342
3163 CATATCCGGTCAAGAAAGACGCTTCGCAACAACTCGGCAAGCAGA 3212
1342 ySerIleAsnAlaIleAsnValThrLeuAsnThrThrGlyThrIleuThrT 1359
3213 AGCCAAAAGACGCGGCAAAAAGACAGCGCGCAAGCCTTG 3253
1359 hrValIysGlySerAsnIleAsnAlaThrSerGlyThrIleu 1372
seq_name: /cgn2_6/prodata/1/laa/5B_COMB.pep:US-08-617-697-2
seq_documentation_block:
? Sequence 2, Application US/08617697
? Patent No. 5977336
? GENERAL INFORMATION:
? APPLICANT: Barenkamp, Stephen J
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Shoemaker and Mattare, Ltd.
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/617,697
? FILING DATE: 01-Apr-1996
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/302,832
? FILING DATE: 05-Oct-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US PCT/US93/02166
? FILING DATE: 16-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Berkstresser, Jerry W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-557
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0810
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1536 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)

```

US-08-617-697-2

Alignment_scores:

Quality: 272.50 Length: 1182
Ratio: 0.452 Gaps: 60
Percent Similarity: 51.015 Percent Identity: 20.897

Alignment_block:

US-09-303-518D-649 x US-08-617-697-2

Align seg 1/1 to: US-08-617-697-2 from: 1 to: 1536

```
127 GGACACACTATTTCGGCATCACTACCAATACTATGCGACTTGGCGCA 176
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 GYGLIUTHTYTLLEUGLY.....GLYAS 356
177 AAATAAAGGCAAGTTTGCAGTCGGGCGAAAGATATTGAGGTTTACACA 226
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 PGLUARGLYGLU.....GLYASANGLYILEGLINLEUALALAYSL 370
227 AAAAAGGAGATTGTCGCAATCAATGACAAAGCCCGCATGATGAT 276
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 YS.....ThrsLeuGluYsglySerThrIleasn 380
277 TTTTCTGTGTGTCGCTAACGGCGTGGCGCATTTGGTGGCGATCAATA 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 ValSerGlysglysglysglyArgAlaIleValTrpGlyAspIleal 397
327 TATGTG.....ACGCTGGCACATA 346
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 aleuIleAspIlyAsnIleasnIagInglySerGlyAspIleAlaIlyst 414
347 ACGGCGGC..... 354
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleIlyAspAsn 430
355 .....TATAACAAGTTGATT 371
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
431 AlaIleValAspAlaIlysglyTrpLeuLeuAspPheAspValSerIl 447
372 TGTGCGCGAA.....GGAGAATATCCCGATCAACATCGTTTACTTAT 414
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 eAsnAlaGluThrAlaGlyIlySerAsnThrSerGluAspAspGlyIly 464
415 .....AAATTTGTAAGCAAGATTAATTAATAAGCA 444
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 hrGlySerGlyAsnSerAlaSerThrProIlyAsnIlysglyIlysthr 480
445 GGGACTAAAGCCATCTTATGGCGGATTAATCAATATGCCGCGTTGCA 494
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 ThrLeuThrAsnThrThrLeuGluSerIleLeuIlysglyIlyThrPhe 497
495 TAAATTTGTACAGATGCAGAACCTGTTGAAATGACAGCATTAATGAGAG 544
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 alAsnIleThrAlaAsnGlnArg.....IleTyVal 507
545 GCGCGAATAATATGATCAAAATTAATTAACCTGACCTGTTCTGATGGG 594
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGlu 524
595 GCAAGGCAAGC.....AATATTGGCGATC 617
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 yArgSerGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541
618 TGATGAGATGAGCCCAATAAC.....GGAAAGTT 649
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 spThrArgGlyAlaAsnLeuThrIleTySerGlyIlyTrpValAspVal 557
650 CATATCATATTGCAAGTGCATTTCTGCGCTGTGGTGCATTAACCTTT 699
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 HisIlyAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573
```

```
700 GCACAAAT.....GATCA.....GCTGG 719
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
574 LysGlnAspIleAlaPheGluIlysglySerAsnGlnValIleThrGlyI 590
720 TGGCAGCTCAACTTAGTAGTGAATAAATTAACATAGCCCATATGCTT 769
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 nGlyThrIleThrSerGlyAsnGlnIly.....GlyP 601
770 TTTTACCAAGAGGCTCATTTGGCGACAGTGCACCAATGTTTATC 819
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuInPhe 617
820 TATGATGCCCAAAAGCAAGTGTAAATTAATGAGGTATGCAACCGG 869
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
618 ThrThrIlyArgThrAsnIlystYrAlaIleThrAsnIlyPheGluGlyTh 634
870 CAACCCCTTATAGCAAAAGCAAT...GGCTTCCAGCTGCTTGTAAAG 916
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
634 rIleAsnIleSerGlyIlyValAsnIleSerMetValLeuProIlyAsnG 651
917 ATTGCTTCTATGATGAATCTTGTCTGGAGATACCAT.....TCAGTA 960
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 lIserGlyIlyAspIly...PheIlysglyArgThrTyThrAsnLeuThr 666
961 TTCTACGAACACAGTCGCAAAATGGAATACTCTTTTAACGAGATATAA 1010
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 SerLeuAsnValSerGluSerGlyIlyPheAsnLeuThrIleAspSerAr 683
1011 TGGCACA.....GGAATAATCAATGCCAATGCAATGCAATCTC 1051
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 gGlySerAspSerAlaGlyThrLeuThrGlnProIlyAsnLeuAsnGlyI 700
1052 TGCGCT...AATGATTAAACACAGACCGTTCATTTGTTAATGTTCT 1098
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 lIserPheAsnIlyAspThrThrPheAsnValGluArg...AsnAlaArg 715
1099 TTATCGAGACAGACAGAACCTGTTATCATGCTGAGGTGTCGCA 1148
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 ValAsnPheAspIleIlysalProIle.....GlyIleAs 727
1149 CAGTTATGACCCACAGCTGAATATGAGAAATATTTCTTATTGACG 1198
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 nIlystYr.....SerSerLeuAsnTyAlaSer 736
1199 AAGAAAAGCGCAATTGATCTTACAGCAATCAATCAAGGCTGCA 1248
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
737 .....PheAsnGlyAsnIleSerValSerIlyGly 746
1249 GATTTATATTTCAGAGAGATTTAAGCTGCTGCTGAAATTAAGAAAC 1298
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 Gly.....SerValAspPheThrLeuAlaIleSerSerAsnVal 760
1299 TTGGCAAGCGCGGCGTT.....CATATCAGTG 1327
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 l...GlnThrProGlyValValIleAsnSerIlystYrPheAsnValSer 776
1328 AAGACAGTACCGTTACTTGAAGTAACGCG..... 1359
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
776 hrGlySerSerLeuArgPheIlysthrSerGlySerThrIlysthrGlyPhe 792
1360 ...GTGGCAACGAC...CGCTGTCCAATTCGCAAGACGACGCTGCA 1403
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
793 SerIleGluIlyAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrIle 809
1404 CGTTCAGGCCAAAGGGAAC.....C 1426
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
809 uLeuGlnValGluGlyThrAspGlyMetIleGlyIlysglyIleValAla 826
1427 AAGGTCGATCAGCGTGGCGAGCTAGACATATTTGGATCAGCAGCA 1476
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
826 yIlyAsnIleThrPheGluGlyGlyAsnIleThrPhe..... 838
1477 GACGATAAAGCAAAACAGCTTTTGTGTAATGCGCTTGTCAGCG 1526
```

```

839 .....GlySerArgLysAlaValThrGluLe.....GluG 849
1527 CAGGGTACGGTCACTGAATGCCGATTAATCACTTCAACCCGACAAAC 1576
1576 .....GlySerArgLysAlaValThrGluLe.....GluG 849
849 yAsnValThrIleAsnAsnAlaAsn..... 858
1577 TCATTTTCGGCTTTCGGGGGAGCGGATTTGGATTAAACGGGCAATGCGCT 1626
859 .....ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCACCGCTATCAAAATACCGATGAAGGGCGGATGTTCTCAACCA 1676
867 AspnAsnIstGlnLysProLeuThrIleLysLysAspValIleIleAsnse 883
1677 CAATCAA.....GACAAAGAAATCCACCTTACCATTCAGGCAATGAAG 1720
883 rGlyAsnLeuThrAlaGlyLysAlaValIleAlaGlyLysLeuThr 900
1721 ATATTGCTACACCGGCAAT..... 1740
900 hrValGlySerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
1741 .....AACACAGCTTGATGACCAAAAGAAATGCGCTA 1775
917 ValGlyLysLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLys 933
1776 CAACGGT...TGTTTGGCGAAGAAAGATACGACCAAAACGAGGGCGG 1822
933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....AsnL 947
1823 TCAACCTGTTTACCAAGCCGCGGAGAGACCGACCTGCTGCTTCC 1872.
947 euserIleThrThrAsnSerSerThrTyrrGlyThrIleIleSerGly 963
1873 GCGGGAACAATTTAAAGGCAATCAACGCAAAACAAAGGCAACTGT 1922
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTGCGGAGACACACACCGGCGCTCAATCATTTAAAGGACAT 1972
977 .....GluGlySerAspThrGlu.....MetGlnIleGlyLysP 989
1973 GGTGCAAAAGAGGGCATTCCTCGCGGAAATCGTGGGACAAACGAC 2022
989 alSerGlnLysGlu.....GlyAsnLeuThrIleSerSerAsp 1001
2023 TGGTCAACGGCACATTTAAAGGCAAACTTCAATTAAGGCGGACA 2072
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGly.. 1014
2073 GCGGCTGTTCCCGCAATCTTGCCAAAAGTGAAGGCGCATGCGATTGA 2122
1015 .....ValAspGlyGluAsnSerAspSerAspAlaI 1025
2123 GCAATCAACGCCCAAGCATTTTGGTGTGCGACCGCATCAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACAGTTCGAGCTGAGCGGTCTGACAATTTGTGCAAAAAC 2222
1032 IleLysThrLys.....GluLeuLys 1038
2223 CATTAACGAGATTAAGTATGCTTCATTAAGTGAAGCGACATGACG 2272
1038 sleuthrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGlnIleThr 1055
2273 GCAATGTCGATCTTGCATCAGCTCATTTAAATCTCACAGGGCTTGGC 2322
1055 lalys...AspGlySerAsp.....Leu 1061
2323 ACATCAACGGCAATCTTATGCAAAATGCGGATACAGTATATACATGAC 2372

```

```

1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078
2373 CCACAACGCCACCCAAACGCGCAACCTTAGCTGCTGATGCGCAATGCCAAG 2422
1078 rPheAsnGlnLysAspSerLysIleSerAlaAspLys..... 1091
2423 CAACATTTAAATCAACCCACATTTAAAGGCAACACATCGCTTGGGCAAT 2472
1092 .....HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105
2473 GCTCATTTAAATCAAGGACGACCGCTACAAAACGCGACGCTGACGCT 2522
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAlaGlyLeuThrII 1121
2523 TTCCGCAACGCTAAGGCAACGTAAGCATTCGCACTCAACGTAATG 2572
1121 e.....AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132
2573 TCTCCCTAGCCGATTAAGCATATTCATTTTGAAGACGCGCTTACC 2622
1132 IeThr...SerHisLysAlaVal.....SerIleSerAlaThrSer 1144
2623 GACAAATCAAGCGCGCGCAAGATACGCAATTCATTAATAA.....GA 2666
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161
2667 CAGCAATGAGACGCTCCGTCAGCGACGCAATTAAGCAATTTAAACCTTG 2716
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyLysIleGlySer 1178
2717 ACAAGCCACCATTTCACTCAATTCGCGCTATCGCACAGATGGCGGAGG 2766
1178 eSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194
2767 GCGCAACCGCGCATGTCGACAGATGCGCGCGCGCTGCGCGCGCTTC 2816
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211
2817 GCGCGTTCCTTATTAATCCGTTACACGCGCACTTCGCTAGAAATCCGCT 2866
1211 rThrLeuAlaGlySerThrIleLysGlyThrCluserValThrThrSers 1228
2867 TCAACACGCTGACGCTAAGGCAATTAAGGCAATTAAGGCAATTAAGGCA 2910
1228 eArgIleSerGlyAspIleGlyLysThrIleSerGly...GlyThrValGlu 1243
2911 TTCGCTTATGTCGAACTTCGCTACGCGACGACGCAATTAAGT 2960
1244 ValLysAlaThrThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259
2961 GCGGAAAGTTCGAAGCACTTACACCTTGCGGCTC..... 2997
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276
2998 .....AACAAATCCGCGCAACGCTGCAAGGCTGCAACCAATG 3036
1276 LysThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292
3037 ACGGTA.....GTGAGAGGAAAGAACAA 3062
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309
3063 ACCGCTGCCGAACCTTAATTCACCTGCAAAACGACAGCATGCGATG 3112
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326
3113 CCGGCGCGTGGCTTACCACTCATCCGCAAGACGCGAATTCGCGCTG 3162
1326 erAlaLysGlyGlnValAlaAsnLeuSerAlaGlnAspGlySerValAlaG 1342
3163 CATATCCGCTCAAGAACAGACTTTCGCAAAATGCGGACGCGACAGA 3212
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThr 1359

```

3213 ACCCAAAAACAGCGGAAAAAGACACGCGAAGCCTTG 3253
1359 hVallysylSerAsnIleAsnAlaThrSerGlyThrLeu 1372

seq_name: /cgr2_6/ptodata/1/iaa/6b_COMB.pep:US-08-719-641-2

seq_documentation_block:

Sequence 2, Application US/08719641

Patent No. 6218141

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Shoemaker and Matzare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Berkstesser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-625

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

US-08-719-641-2

alignment_scores:

Quality: 272.50 Length: 1182

Ratio: 0.452 Gaps: 60

Percent Similarity: 51.015 Percent Identity: 20.897

alignment_block:

US-09-303-518D-649 x US-08-719-641-2 ..

Align seg 1/1 to: US-08-719-641-2 from: 1 to: 1536

127 GCACACTTATTTGGCATCACTACCAATACATCGGACTTTGGCA 176
|||||
349 GlycylThrTyrluGly.....GlyAs 356
177 AATAAAGCAAGTTTGCATCGGGGCAAGATATTGAGTTACACA 226

356 pGluArgGlyGlu.....GlyAsnGlyIleGlnLeuAlaLysL 370
227 AAAAAAGGAGTTGTGCGCAATCAATCAAAAGCCCGCATGATTGAT 276
|||
370 ys.....ThrSerLeuGluLysGlySerThrIleAsn 380
277 TTTTCTGTGTGTCGCGTACGGCGGCGCATTTGTTGGCGCATGAATA 326
|||
381 ValSerGlyLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397
327 TATTTGTC.....ACGTCGCACATA 346
397 aleuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlLys 414
347 ACGGCGGC..... 354
414 hrglyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430
355TATACACAGCTGATT 371
431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspValSerI 447
372 TGGTGGCGAA.....GGAAGAAATCCGATCAACATCGTTTACTTAT 414
|||||
447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGlyTr 464
415AAATGTGAAACGGAAATATTAAACA 444
464 hrglySerGlyAsnSerAlaSerThrProLysArgAsnLysGlyLys 480
445 GGCACATAAGGCCATCTTATGCGGCGGATTTATCATATCGCGTTTGA 494
|||
481 ThrLeuThrAsnThrThrLeuLysSerIleLeuLysGlyThrPhe 497
495 TAAATTTGTCACAGATGCACAGACCTGTGAATGACAGTATATGAGAT 544
|||||
497 alaSerIleThrAlaAsnGlnArg.....IleYrVal 507
545 GGGGGAATATATGCAATCAAAATATATACCGTCGCTGTTGATGG 594
508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGlu 524
595 GCAGGCGAGC.....AATATTGGCGATC 617
|||||
524 yArgSerGlyGlyValGluLysAsnAspIleThrThrGlyAspA 541
618 TGATGAAGATGACCCCAATATACC.....GCGAAGAT 649
541 spThrArgGlyAlaAsnLeuThrIleYrSerGlyLysTrpValAspVal 557
650 CATATCATATTTGCAAGTCGCTATTTGCGTCGTCGTCGTCGTCGTC 699
|||
558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573
700 GCACAAAT.....GGATCA.....GGTG 719
|||
574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGly 590
720 TGGCAGACGTCACCTTAGTAGTGAATAATTAACATAGCCCATATGCT 769
|||||
590 nglyThrIleThrSerGlyAsnGlnLys.....GlyP 601
770 TTTTACCAACAGAGGCTCATTTGGGACAGTGGCTCACCAATGTTTATC 819
|||
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617
820 TATGATGCCCAAAAGCAAAAGTGTAAATTAATGAGGTATGCAACGG 869
|||
618 ThrThrLysArgThrAsnLysTrpAlaIleThrAsnLysPheGluGly 634
870 CAACCCCTATATAGCAAAAGCAAT...GCCTTCACGCTGCTTCGTAAG 916
|||||

634 rleuasnlleSerIgLysValasnlleSerMetValleuProLysAsnG 651
917 ATTGGTTCTATGATGAATCTTGTGAGATACCCAT.....TAGTA 960
::: |||||::: ||| ||| ||:::::: :
651 luserGIlyTyrAspLys...PheLysGIlyArGhTyrTTPasnleuthr 666
961 TTTCAGACACCGCTCAAAATGGGAATACTCTTTTAACGACGATATAA 1010
::: |||||::: |||||::: |||||::: |||||::: :
667 SerleuasnlleSerGIlySerGIlyLupheasnleuthrIleAspSerAr 683
1011 TGGCACA.....GAAAAATCAATGCCAACATGAACACAAATCTC 1051
|||::: |||::: |||::: |||::: |||::: :
683 gGIlySerAspSerIleAgIlyThrleuthrGlnProTyrAsnleuasnllyI 700
1052 TGCCCT...AATGATTTAAACAGCAACCGCTCAATGTTAATGTTCT 1098
|||::: |||::: |||::: |||::: |||::: :
700 leSerPheasnLysAspThrThrPheasnValGIlyArg...AsnAlaArg 715
1099 TTATCCGACGACGACGACGACCTGTTTATCATGCTCAGGTGGGTGCA 1148
::: |||||::: |||||::: |||||::: |||||::: :
716 ValasnPheAspIleLysAlaProIle.....GIlyIleAs 727
1149 CAGTTATGACCCGACGACTGAAATATGAGAAAATTTCTTATTACG 1198
|||::: |||::: |||::: |||::: |||::: :
727 nLysTyr.....SerSerleuasnltyrAlaSer. 736
1199 AAGGAAAAGGCGAATGTACTACTACGACGACATCAATCAAGGTGCGGA 1248
737PheasnGIlyAsnIleSerValSerGIlyGIly 746
1249 GGATTTATATTTCAGAGGATTTTACGGTCTCGCCTGAAAATAAGCAAC 1298
|||::: |||||::: |||||::: |||||::: :
747 Gly.....SerValAspPheThrleuLeuAlaSerSerAsnVal 760
1299 TTGGCAGGCGCGCGCGCT.....CATTCAGTG 1327
::: |||::: |||||::: |||||::: |||||::: :
760 L...GlnThrProGIlyValIleAsnSerLysTyrPheasnValSerT 776
1328 AAGACGACGCTTACTTGAAGATAAGCAAGC.....1359
|||::: |||||::: |||||::: |||||::: :
776 hrGIlySerSerleuasnrPheLysThrSerGIlySerThrLysThrGIlyPhe 792
::: |||||::: |||||::: |||||::: |||||::: :
1360 ..GTGGCAAGCAGC...CGCCTGTCCAAATCGGCAAGGCGACGTGCA 1403
::: |||||::: |||||::: |||||::: |||||::: :
793 SerIleGIlyLysAspLeuThrleuasnlleThrGIlyGIlyAsnIleThrIle 809
1404 CGTTCAGACCCAAAGGGGAAC.....C 1426
::: |||||::: |||||::: |||||::: |||||::: :
809 uLeuGIlyValGIlyGIlyThrAspGIlyMetIleGIlyLysGIlyIleValAlaL 826
1427 AAGGCTCGATCAGCGTGGCGCGACGATCATTTTGATCAGCAGGCA 1476
::: |||||::: |||||::: |||||::: |||||::: :
826 yLysAsnIleThrPheGIlyGIlyGIlyAsnIleThrPhe..... 838
1477 GACGATAAGGCAAAACAGCCTTAGTGAATCGGCTTGGTCACCG 1526
|||::: |||||::: |||||::: |||||::: :
839GIlySerArgLysAlaValThrGIlyIle.....GIlyGI 849
1527 CAGGGGTACGGTGCACATGATGCCGATATCAGTTCAACCCGACAAAC 1576
|||::: |||||::: |||||::: |||||::: :
849 yAsnValThrIleAsnAsnAlaAsn..... 858
1577 TCTATTTCGGCTTTCGGCGGACGTTTGATTTAAAGGCGCATTCGCT 1626
::: |||::: |||::: |||::: |||::: :
859ValThrleuIleGIlySerAspPhe 866
1627 TCGTTCACCGCTATTCAAAATACCGATGAAGGGCGATGTGTCACCA 1676
::: |||||::: |||||::: |||||::: |||||::: :
867 AspAsnHisGIlyLysProleuthrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA.....GACAAAGAATCCACCGTTACCATTCAGCAGCATTAAG 1720
::: |||||::: |||||::: |||||::: |||||::: :
883 rGIlyAsnleuthrAlaGIlyGIlyAsnIleValIleAlaGIlyAsnleuthr 900

1721 AATATGCTACAAACCGCAAT..... 1740
::: |||||::: |||||::: |||||::: |||||::: :
900 hrValGIlySerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheasn 916
1741ACACACGCTTGATAGCAAAAGAAATGGCTTA 1775
|||::: |||||::: |||||::: |||||::: :
917 ValGIlyGIlyLeuPheAspAsnLysGIlyAsnSerAsnIleSerIleAlaLys 933
1776 CAACGGT...TGTTGGCAGAAAGATACGCAAAACGAAAGGCGCGC 1822
|||::: |||::: |||::: |||::: |||::: :
933 sGIlyGIlyAlaArgPheLysAspIleAspAsnSerLys.....Asn 947
1823 TCACCTGTTTATCCAGCCCGCGCAGAAAGACCGACCGCTGCTGCTTC 1872
|||::: |||::: |||::: |||::: |||::: :
947 euserIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGIly 963
1873 GCGGACACAAATTTAAGCGCACATCCAGCAACAAACGGCAAACTGT 1922
::: |||||::: |||||::: |||||::: |||||::: :
964 AsnIleThrAsnLysAsnGIlyAspLeuasnlleThrAsn..... 976
1923 TTTCAGCGGCGACACCAACCGCGCTACAAATCATTTAAACGACATT 1972
|||::: |||::: |||::: |||::: |||::: :
977 ...GIlyGIlySerAspThrGlu.....MetIleIleGIlyGIlyAspV 989
1973 GGTGCAAAAAGAGGGGCAATTCCTCGGGGAAATGCTGTGGGACAGCAG 2022
|||::: |||||::: |||||::: |||||::: :
989 alSerGIlyLysGlu.....GIlyAsnleuthrIleSerSerAsp 1001
2023 TGATATCACCGCACATTTAAGCGGAAACTTCCAAATTTAAAGGGGACA 2072
|||::: |||||::: |||||::: |||||::: :
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGIly.. 1014
2073 GCGGGTGTTTCCCGCAATGTTGCCAAAGTGAAGGCGATTGGCATTTGA 2122
|||::: |||::: |||::: |||::: |||::: :
1015ValAspGIlyGIlyLysAsnSerSerAspAlaIle 1025
2123 GCATTCACGCCGACAGCAGTTTGTGTGCGCACCGCATCAAGCCACACA 2172
|||::: |||||::: |||||::: |||||::: :
1025 hrAsnAsnAlaAsnleu.....Thr 1031
2173 ATCTGTACACGTTGCGACGTGACGGGTCTGACAAATGTGTGGAATAAAC 2222
|||::: |||::: |||::: |||::: |||::: :
1032 IleLysThrLys.....GluLeuLys 1038
2223 CATTCACGACGATTAAGATGATGCTTCAATGCTAAGACCGACATCAGCG 2272
|||::: |||||::: |||||::: |||||::: :
1038 sleuthrGlnAspLeuasnlleSerGIlyPheasnLysAlaGIlyIleThrA 1055
2273 GCAATGTGATCTTGCCGATCAAGCTCATTTAATCTACAGAGGCTTGCC 2322
|||::: |||::: |||::: |||::: |||::: :
1055 lAlys...AspGIlySerAsp.....Leu 1061
2323 ACACTCACGCAATCTTGTGCAATGGCGATACAGCTTATACAGTCAG 2372
|||::: |||||::: |||||::: |||||::: :
1062 ThrIleGIlyAsnThrAsnSerAlaAspGIlyThrAsnAlaLysLysValThr 1078
|||::: |||||::: |||||::: |||||::: :
2373 CGACAAGCGCACCAAAACGCAACCTTAGCCTGTGGGCAATGCCCAAG 2422
|||::: |||::: |||::: |||::: |||::: :
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGIly..... 1091
2423 CAACATTTAATCAAGCCACATTTAAGCGCACATCGGCTTGGGCGCAT 2472
|||::: |||||::: |||||::: |||||::: :
1092HisLysValThrleuHisSerLysValGIlyThrSerGIlySer 1105
2473 GCTTCATTAATCTAAGCAGCAGCGCTACAAACGCGAGCTGACGCT 2522
|||::: |||::: |||::: |||::: |||::: :
1106 AsnAsn...AsnThrGlnAspSerSerAspAsnAlaGIlyLeuThrIle 1121
2523 TTCCGCAACGCTTAAGCAAGATTAAGCCATTCCGCACTCAACGGTAATG 2572
::: |||||::: |||||::: |||||::: |||||::: :
1121 e.....AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132

```

2573 TCTCCCTGAGCGGATAGGACGATATTCATTTTGAAGACGCGCTTACC 2622
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1132 l e t h r . . . . . S e r i e s l y s a l a v a l . . . . . S e r i e s l a t h r s e r 1144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2623 GGACAAATACAGGGGGGCAAGGATACGGCATTCACCTTAAAAA.....GA 2666
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1145 G l y t h r l e u t h r l e u t h r l y s t h r g l y t h r l e u s n a l a t h r t h r g l y s 1161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2667 CAGCGAATGAGCGCTGCGCGACGACGAGATTCAGCAATTTAAACCTG 2716
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1161 n v a l g l u t l e t h r l a g l i n t h r g l y s e r i l e u g l y g l y l e u g l u s e r s 1178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2717 ACAACGCCACCATTCACACTCAATTCGCGCTTACGCCACGATGCGGACGG 2766
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1178 e r s e r g l y s e r v a l t h r l e u t h r l a t h r g l u g l y a l a l e u a l a v a l s e r 1194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2767 GCGCAACCGGCGAGTCGACAGATCGCGCGCGCGCGCTTCCGCGCTC 2816
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1195 A s n i l e s e r g l y a s n t h r v a l t h r v a l t h r a l a s n s e r g l y a l a l e u t h 1211
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2817 GCGCGCTTCCCTATTATTCCTTACACGCGCAACTTCGTTACGATCCGCTT 2866
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1211 r t h r l e u a l e u l y s e r t h r l e u t h r g l y t h r g l u s e r v a l t h r t h r s e r s 1228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2867 TCACACGCTGACGCTAAACGCAATTTGAACGGTCAAGGACAA..... 2910
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1228 e r g l n s e r g l y a s p l e g l y g l y t h r l e s e r g l y . . . g l y t h r v a l g l u 1243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2911 TTCGCGCTTATGTGGAACCTTTCGGCTACGCGACGCAACAATTTGAAGCT 2960
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1244 V a l l y s a l a t h r g l u s e r l e u t h r t h r g l n s e r s e r l y s t h r l y s . . 1259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2961 GCGGAAAGTTCCGAGGACCTTACACTTCGCGCTG..... 2997
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1260 . A l e t h r t h r g l y g l u a l a s n v a l t h r s e r a l a t h r g l y t h r l e u g l y 1276
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2998 .....ACAAATACGCGGCAAGACCTGCAACCTTCGCAACAATG 3036
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1276 l y t h r l e s e r g l y a s n t h r v a l a s n v a l t h r a l a s n a g l y a s p l e u 1292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3037 ACGGTA.....GTGGAAGGAAAGCAACAA 3062
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1293 T h r v a l g l y a s n g l y a l a g l i l e a s n a l a t h r g l u g l y a l a a l a t h r l e 1309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3063 ACCGCTTCGCAAAACCTTAAATTCACCTGCAAAACGACGCTGCATG 3112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1309 u t h r t h r s e r s e r g l y s l e u t h r t h r g l u a l a s e r s e r i s l e t h r s 1326
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3113 CCGGCGCTGCGCTTACCACTCATTCGCAAGAGCGGAGTTCGCGCTG 3162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1326 e r . A l a l y s g l y g l n v a l a s n l e u s e r a l a g l n a s p l y s e r v a l a l a g l 1342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3163 CATATCCGCTGCAAGAAACAGAGCTTCGCAACAACCTCGGCAAGCAGA 3212
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1342 y s e r i l e u s n a l a l a s n v a l t h r l e u a s n t h r t h r g l y t h r l e u t h r t 1359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3313 AGCCAAAAACAGCGCGGAAAAACAGCAAGCGCAAGCCTTG 3253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1359 h r v a l l y s g l y s e r a s n l e a s n l a t h r s e r g l y t h r l e u 1372
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-169-927-2
seq_documentation_block:
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Chang, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both

```

```

? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Counsel, Naval Medical R & D Command
? STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
? CITY: Bethesda
? STATE: MD
? COUNTRY: USA
? ZIP: 20889-5606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/169,927
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/742,128
? FILING DATE: 08/09/91
? ATTORNEY/AGENT INFORMATION:
? NAME: Spevack, A. David
? REGISTRATION NUMBER: 24,743
? REFERENCE/DOCKET NUMBER: 75,976
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 295-6759
? TELEFAX: (301) 295-1022
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1612 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-169-927-2

```

```

alignment_scores:
  Quality: 241.00      Length: 1485
  Ratio: 0.361         Gaps: 75
  Percent Similarity: 44.916   Percent Identity: 19.259

```

```

alignment_block:
US-09-303-518D-649 x US-08-169-927-2 ..

```

```

Align seg 1/1 to: US-08-169-927-2 from: 1 to: 1612

```

```

634 AATAACCGGAAAGTTCATATCATATTCGCAAGCTGCTATTCCTGGCTCGT 683
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 A n s e r a r g s e r s e r s e r t h r i s l e u a l s e r . . . . . 170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
684 TGGTGCATATTCCTTGCACAAATGAGATCAGGTGCTGTCACAGTCACACT 733
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
171 .....A n s e r L y s i l e a l a s n g l y . . . g l y s n g l y l l e u a s n i 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
734 TA.....GGTAGTGAATAAATTTAAACATAGCCCATATGTTTATACCA 777
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 l e t h r a s n g l y p h e l l e g l n v a l s e r a s p a s n t h r p h e a l a g l y l l e u y s 200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
778 ACAGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
201 T h r . . . . . l l e a s n l l e a s p a s p c y s g l n g l y l e u m e t p h e a s n s e r t h r p r 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
828 CCAAAAGCAAAAGTGCTA.....A 847
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
216 o a s p a l a l a s n t h r l e u a s n l e u g l n a l a g l y l y s n t h r l l e a s n p 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
848 TTAATGGGCTATTGCAAACGCGCAACCCCTATATAGAAAGCAATGCG 897
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
233 h e a s n o l y l l e a s p l y t h r g l y s l e u a l s e r l y s a s n g l y 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
898 .....TTCAGCTGCTTCTGTAAGCTTGGTCTTATGATGAAT 935
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
250 A l a l a t h r g l u p h e a s n v a l t h r g l y . T h r l e u g l y g l y s n l e u l y s g 266

```

[illegible][illegible]

[illegible]

```

3938 GCGTGTGCATTAAGGCAATTCAGCAGATACCGCGCGGTTTCGGCGA 3987
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1277 Ala.....ValThrThrValAsnLeuAsnAspThrGlnIsthrGlnAs 1291
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3988 TTCGCAATCGAAGCCGACATCGCGGCGGCGCTATTTCGCCAAAAGC 4037
      - - - - - :||| :||| :||| :||| :||| :||| :||| :||| :|||
1291 pLeuLeuSerAsnArg..LeuGlyThrLeuAlaGlyThrLeuSerAsnAla.. 1306
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4038 GGATTACCGGTACGAAACGTCATATCGCCACCGCCCGGCTTCGCATTC 4087
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1307 .....GluThrSerAspAlaAla.....GlySerAlaThrG 1317
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4088 ACCGCTACCGCGCGGCAATTAAGCAGATTATTCATTAACCGCGCA 4137
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1317 LysAlaValSerSerGlyAspGlnAlaGluValSerTyr..... 1329
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4138 CACATTTTCATCAGCGCTTATTGAGCTGTCTGCTATACGATGCGCTTC 4187
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1330 GLyValThrPalalysProphETyrAsnIleAlaGluGlnAspIstysG 1346
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4188 GGGC.....AAGTCCGACACCGC..... 4206
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1346 YGlyIleAlaGlyTyrIstysAlaIsthrThrGlyValAlaGlyLeuAl 1363
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4207 .....GTC 4209
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1363 sPThrLeuAlaSerAspAsnLeuMetIleGlyAlaAlaIleGlyIleThr 1379
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4210 AATACCCCGCTATTGGCTCAGAGATTTCGCAAAACCGCGAGTGGCAATG 4259
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1380 LysThrAspIleLysHisGlnAspTyrIstys..... 1390
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4260 GGGGTAAACCGCGAATCAAGGTTTCAGCGTGTCCCTCCAGCGTGC 4308
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1391 .GlyAspIsthrAspIleAsnGlyLeuSerPheSerLeuTyrGlySer 1406
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /cgn2_6/prodata/1/1aa/6B.COMB.pep:US-09-268-347-36

```

```

seq_documentation_block:
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: prt
; ORGANISM: Haemophilus influenzae
; US-09-268-347-36

```

```

alignment_scores:
  Quality: 240.50      Length: 1553
  Ratio: 0.357        Gaps: 80
  Percent Similarity: 43.400  Percent Identity: 19.446

```

```

alignment_block:
US-09-303-518D-649 x US-09-268-347-36 ..

```

```

Align seg 1/1 to: US-09-268-347-36 from: 1 to: 2411

```

```

202 GCGAAGATATTGAGTTTACACAAAAAGGAGTTGTC..... 243
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
444 SerIstysAspGlnGluVal.....LysAlaGlyGluThrValAlaThrPhe 458
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
244 ....GCGAATTCATG...ACAAAGCCCGATGATGATTTCCTGTGG 286
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

458 sAlaGlyLysAsnLeuLysValLysGlnAspGlyAlaAsnPheThrTyrS 475
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
287 TGTCCGCTAACGGCGGTGGCGGATGTGTGGC..... 318
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
475 erLeuGlnAspAlaLeuThrThrGlyLeuThrSerIleThrLeuGlyGlyThr 491
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
319 .....GATCATATAT 329
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
492 ThrAsnGlyGlyAsnAspAlaLysThrValIleAsnLysAspGlyLeuTh 508
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
330 TGTGAGCGGTGCACATTAAGCGGCGCTATACACAGCTGATTTCGTGGCG 379
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
508 rIleThrProAlaGlyAsnGly.....GlyThrT 518
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
380 AAGGAAGAATCCCGATCATCATCGTTTACTTAAATATGTAACGG 429
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
518 hrGlyThrAsn.....ThrIleSerValThrLys... 527
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
430 AATTAATTAAGACAGGACGCTAAGCCATCCTTATGCGCGCATTAACA 479
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
528 AspGlyIleLysAlaGlyAsnLysAla..... 536
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
480 TATGCGCGCTTGCATTAATTTGTCAAGATGCAGAACCTGTGAATGA 529
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
537 .....IleT 538
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
530 CCAGTTATATGATGGCGGAATATATCATCAAAATTAATACCTGAC 579
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 hrAsnValAlaSerGlyLeuArgAlaTyrAspAspAlaAsnPh..... 552
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
580 CGTGTTCGATTGGGCGAGGACGCAATATTGGCATGTGATGAAGATGA 629
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
553 .....AspVa 554
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
630 GCCAATAAC.....CGGAAAGTTCAATCATATTGCAAGTGGCTAT 673
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
554 LLeuAsnAsnSerAlaThrAspLeuAsnArgHisValGlnAspAlaTyrL 571
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
674 CTGGCTGCTGTGGCAATACCTTTGCACAAATGATCAGTGTGGC 723
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
571 ys.....Gly 572
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
724 ACAGTCACTTGTAAGTGAATAAATTAACTAGCCCATATGTTT 773
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
573 LeuLeuAsnLeuAsnGlnLysAsnAlaAsnLysGlnProLeuValThrAs 589
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
774 ACCAACAGGAGGCTCATTTGGCGACAGTGGTCACCAATGTTATCATAG 823
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
589 pSerThrAlaAlaThrValGly.....A 597
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
824 ATGCCCAAAAGCAAAAGTGTAAAT.....AATGGGTATTCGAA 864
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
597 sPLeuThrGlyLeuGlyTyrPValValSerThrLysAsnGlyThrLysGlu 613
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
865 ACGGCAACCCCTATATAGAAAAAGCAATGCTTCGACGTCGTTCGTA 914
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
614 GluSerAsn.....GlnValLysGlnAl 621
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
915 AGATTGTTCTATGATGAATCTTGTGAGATACCCATCAGATTCT 964
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
621 asp.....GluValLeuPheThrGlyAlaGlyAlaAlaThrValT 635
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
965 ACGAACACAGTCAAAATGGGAATACTCTTT..... 996
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
635 hrSerLysSerGlnAsnGlyLysHisThrIleThrValSerValAlaGlu 651
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
996 ..... 996
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
652 ThrLysAlaAspSerGlyLeuGlnLysAspGlyAspThrIleLysLeuLys 668
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
997 .....AACGACGAT.....AATAATGCA 1015
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
668 sValAspAsnGlnAsnThrAspAsnValLeuThrValGlyAsnAsnGlyT 685
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```
1213 LuLeIeThrAsnValLysSerAlaLeuLysThrLysAspThrGln... 1228
      ::::: ||| ::::| |||:::
2446 AACGGCAACACATCGCTTCGGGCAATGCTTCATTATATTAAGCCACCA 2495
      |||||:::
1229 ....AsnThrAlaGlyAlaThrGlnProAlaAlaAsnThrAlaGly 1243
      :::::
2496 CGCGCTCAAAAC.....GGCAGCTGA 2518
      ||| |||:::
1243 LAlaLysGlnSpleuValAspleuThrLysProAlaThrGlyAlaLag 1260
      |||:::
2519 CGCTTCCGGCAACGCTTAAGCAACGTAAAGCATTCGCCATCAACGCT 2568
      :::::
1260 LysAsnGlyAlaAspAlaLysAlaProAspThrThrAlaLthrValGly 1276
      :::::
2569 AATGTCCTCCCTAGCCGATTAAGGCGATTCATTTTCAAAAGCAGCCGCT 2618
      :::::
1277 Aspleu.....ArgCl 1280
      :::::
2619 TACCGCAACATTCAGCGGCAAGATAGCGATTCACCTTAATAAGACA 2668
      ||| :::::
1280 YLeuGlyThrValLeuSerAlaLysLysThrAlaAspLthrGlnAspL 1297
      :::::
2669 GCGAATGG.....ACGCTGCCGTACGACGCAAGATTAGCAATTAAAC 2712
      :::::
1297 YsgLuphEniAlaAlaValLysAsnAlaAsnGluValGluPheValGly 1313
      :::::
2713 CTGACCAACGCCACCATTTACA....CTCAATTCCGCTATGCGCCG.. 2754
      :::::
1314 LysAsnGlyAlaLthrValSerAlaLysThrAspAsnGlyLysHisTh 1330
      :::::
2755 .....GATCGCGCAGGGGCGCAACCGCGATGCG..... 2784
      |||||:::
1330 rValThrLthrSpleuAlaLagLunAlaLysValGlyAspGlyLeuGluLys 1347
      :::::
2785 ..ACAGATGG.....CCGCGCCCGCTTCGCGCGCTTCGCGCGCTTC 2826
      |||||:::
1347 sPrThrAspGlyLysAlaLysLysValAspAsnThrAspGlyAsnAsn 1363
      :::::
2827 CTATTATCCGTT.....ACACGCCCACTCGGTA...GAATCCGCTT 2867
      |||||:::
1364 LeuLthrValAspAlaLthrLysGlyAlaSerValAlaLysGlyLuph 1380
      :::::
2868 CAACACGCTGACG.....GTAACGCGCAAT 2893
      |||||:::
1380 eAsnAlaValThrThrAspAlaLthrThrAlaGlnGlyThrAsnAlaAsnG 1397
      :::::
2894 TGAACGGTCAGGGAACATTCGCTTATGTCGAACCTTCGCTACCGC 2943
      :::::
1397 LuATGgLyLysValValLysGlySerAsnGlyAlaLthrAlaLthrGlu 1413
      :::::
2944 AGCGCAATTTGAAGCTGGGGAAGTTCCGAAGCACTTACACCTTGGC 2993
      :::::
1414 ThrAspLysLysValAla.....ThValGlyAspValAlaLysAl 1428
      :::::
2994 GGTCAACAATACCGCAACGCAACCTGCAAGCCTGCAACATTTGACGGTAG 3043
      :::::
1428 aLLeAsnAspAlaLthr.....PheValLysV 1438
      :::::
3044 TCGAAGGAAGAAGACAAACCGCTGTCGGAACCTTAATTACCCCTG 3093
      |||||:::
3094 CAAGAAGCAACGTCGATGCCGCG.....GGGTGGC 3125
      :::::
1455 AlaAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLysAlaGly 1471
      :::::
3126 TTACCAACTCATCCGCAAGAAGCGCGAG.....TTCCGCTGCATA 3166
      :::::
1471 sAsnLeuLysValLysArgAspGlyLysAsnLthrPheAlaLeuAla 1488
      :::::
3167 AT.....CCGGTCAAGAACAAGAGCTTTCGCAAACTC..... 3201
      |||||:::
1488 snAspLeuSerValLysSerAlaLthrValSerAspLysLeuSerLeuGly 1504
      :::::
3202 .....GCCAGCGAAGAACCAAAACAGCGGGA..... 3231
      :::::
1505 ThrAsnGlyAsnLysValAsnLthrSerAspThrLysGlyLeuAspTh 1521
      :::::
3232 ...AAAGCAACGCGCAAGCCTTGACCGCTGATTCGGCGCGCGCG 3277
      |||||:::
1521 eAlaLysAspSerLysThrGlyAspAspAlaAsnLthrHisLeuAsnGly 1538
      :::::
3278 ATCCGCTCGAAGAACAGAACGCTT.....GCCGAACCGCGCCGCGAG 3321
      |||:::
1538 LeAlaSerThrLeuThrAspThrLeuLeuAsnSerGlyAlaLthrThAsn 1554
      :::::
3322 GCAGCGGGGAATGTCGCATTCATGACGGCGAGAGAAAGAAACG 3371
      |||||:::
1555 LeuGlyGlyAsnGlyLe.....ThrAspAsnGlyLysAsp 1567
      :::::
3372 GGTGCAGCGCGATTAAGACACCGCCTTGCGCAACGCGCAAGCGGAA 3421
      :::::
1567 gAlaAlaSerValLysAspValLeuAsnAlaGlyThrAsnValArgGly 1584
      :::::
3422 CCCGCGCGCTACACACCGCTTCGCCCGCGCGCGCGCGCGGAT 3471
      :::::
1584 aLlYsProAlaSer..... 1588
      :::::
3472 TTGCGCAACTGCAACCCCAACGCGAGCCCAACCGCAGCGCATGTAT 3521
      :::::
1588 ..... 1588
      :::::
3522 CAGCGCTTATGCCAATAGCGTTTGAAT.....GAATTTCCGCCACGC 3565
      |||||:::
1589 .....AlaAsnAsnGlyValGluAsnLthrSpleuAlaLthrT 1602
      :::::
3566 TCAACAGCGTTTCCGCTACGACGACGCAATTAGAC.....CGCGTATT 3609
      :::::
1602 YrAspThrValAspPheValSerGlyAspLysAspThrThrSerValThr 1618
      :::::
3610 GCCGAAGACCGCGCAAC.....GCCGTTG 3635
      :::::
1619 ValGluSerLysAspAsnGlyLysArgThrGluValLysLagLysAla 1635
      :::::
3636 GACACGCGCATCCGGGACACCAACACGCTTCGCAAGTTTCCCGC 3685
      |||||:::
1635 sThrSerValLysLysAsp.....HisAsnGlyLysLeuPheThrG 1649
      :::::
3686 CTTACCGCCCAACAAACCGCAGCTGCGCAATCGATGACG..... 3726
      :::::
1649 Ly...LysGluLeuLysAspAlaAsnAsnGlyValThrValThrGlu 1664
      :::::
3727 .....AAAACCTCGCAGCGGCGCGTC.....GGCATCT 3758
      |||||:::
1665 ThrAspGlyLysAspGluLysAsnGlyLeuValThrAlaLysAlaVal 1681
      :::::
3759 GTTTTCGCAACCGG..... 3774
      :::::
1681 eAspAlaValAsnLysAlaGlyLThrPargValLysThrThrGlyAlaAsnG 1698
      :::::
3775 .....ACGAAACACCTTC 3789
      |||||:::
1698 LysGlnAsnAspAspPheAlaLthrValAlaLaserGlyThrAsnValThrPhe 1714
      :::::
3790 GACGACGGCATCGCAACTCGGACGCGCTTGC.....CACGGCGC 3830
      |||||:::
1715 AlaAspGlyAsnGlyThrThrAlaGluValThrLysAlaAsnAspGlySe 1731
      :::::
3831 CGTTTCGGGCAATACGCATCGACAGGTTTCACTCGGCATCACGCGCG 3880
      :::::
1731 rLlthrValLysTyAsnVal..... 1738
      :::::
3881 GCGCGGTTTATGACGCGGACGCTTTCAGCGGCATC.....GGA 3921
      :::::
1739 .....LysValAlaAspGlyLeuLysLeuAspGly 1748
      :::::
```


1404 CGTTCAAGCAAGGGAACCAAGGCTGCATCAGGTGGGCGAGGTA 1453
1024AsnAspGlyGlyLeuSerIleLysAsnProI 1034
1454 CAGTCATTTTGGATCAGCAGCAGCAGATTAAGCAAAAACAGCCTTT 1503
1034 hrGlySerGluGlnIleGlnValGlyAlaAspGlyValLysPheAla... 1049
1504 AGTGAATCGGCTTGGTCACGCGGAGGGGTACGCTGCACTGAATCCGA 1553
1050LysValAsnAsnAsnGlyValValGlyAlaGlyIleAs 1062
1554 TTAACAGTTCAACCCCGCAAAACTCTATTTCGGCTTTCGGCGC...G 1597
1062 pGlyThrThrArgIleThrArgAspGluIleGlyPheThrGlyThrAsnG 1079
1598 GACGTTTGAT.....TTAAACGGGCAATTCGCTT 1626
1079 LysSerLeuAspLysSerLysProHisLeuSerLysAspGlyIleAsnAla 1095
1627 TCGTTCACCGTATTCAAATACCGATGAGGGCGGATGATTCGAACCA 1676
1096 GLyGlyLysLysIleThrAsnIleGlnSerGlyGluIleGlnAlaAsnSe 1112
1677 CAATCAAGACAAGAAATCCACCGTTACCATTAAGCAATAAA.....G 1720
1112 rHisAsp.....AlaValThrGlyGlyLysIleGlyrAs 1123
1721 ATATTGCTACA..... 1731
1123 sPLeuLysThrGluLeuGluAsnLysIleSerSerThrAlaLysThrAla 1139
1732ACGGCAATTAAC... 1743
1140 GlnAsnSerLeuHisGluPheSerValAlaAspGluGlnLysAsnAsnPh 1156
1744AACAGCTTGATAGCAAAAAGAAATTCCT 1774
1156 eThrValSerAsnProTyrSerSerTyrAspThrSerLysThrSerAspV 1173
1775 ACAACGGTGGTTGGTCGAGAAAGAT...ACGACCAAAACGAAC...GGG 1818
1173 alIeThrPheAlaGlyLysAsnGlyIleThrThrLysValAsnLysGly 1189
1819 CGGCTCAACCTTGTTCACAGCCCGCGAGAGACCGCACCTGCTGCT 1868
1190 ValValAlaGlyValGlyIleAspGlnThrLysGlyLeuThrProLysLe 1206
1869 TTCGGCGGCAAAATTTAAACGGCAAC.....ATCAGCGCAACAA 1909
1206 uThrValGlyLysAsnAsn...AsnGlyLysGlyIleValIleAspSerGlnA 1222
1910 ACGGCAACCTGTTTTCAGCGCGCAGACCAACACCGCCCTACATCAT 1959
1222 snGlyGlnAsnThrIleThrGlyLeuSerAsnThrLeuAla...AsnVal 1237
1960 TTAACAGCACTGTGTCGCAAAAAGAGGCAATTCGCGCGGGAATTCG 2009
1238 ThrAsnAspLysGlySerValArgThr...ThrGluGlnLysAsnIleI 1253
2010 GTGGGAGACACTGATCAACCGCATTTAAACGGGAA..... 2049
1253 eLysAspGluAsp.....LysThrArgAlaAlaSerIleValAspV 1267
2050AACTCCAAATTAAGGC...GGACAGCGGGGTTC 2085
1267 alLeuSerAlaGlyPheAsnLeuGlnLysGlyAlaVal..... 1281
2086 CGCATGTGCGCAAAAGTGAAGCGGATTCGATTGAGCAATACCGCCA 2135
1282AspPhe...ValSerThrTyrAspTh 1289

2136 AGCAGTTTGTGGTGCACCGCATCAAGCCACACATCTGTACAGTT 2185
1289 rValAsnPhe.....AlaAspGlyAsnAlaThrThrAlaLysValThrT 1304
2186 CGGACTGAGCGGGTCTGCACAAATGTTGTGCAAAAACCATTAACGAGAT 2235
1304 rAspAspThrSerLysThrSerLysValValTyrAspValAsnAspAsp 1320
2236 AAGTGAT.....GCTTCATT 2252
1321 ThrThrIleGluValLysAspLysLeuGlyValLysThrThrThrIle 1337
2253 GACTAAGACCGACATCAGCGCAAT..... 2277
1337 uThrSerThrGlyThrGlyAlaAsnLysPheAlaLeuSerAsnGlnAla 1354
2278GTGATCTTCGCGATCAGCGCATTTA 2304
1354 hrGlyAspAlaLeuValLysAlaSerAspIleValAlaHisSerLeuAsn 1370
2305 AATTCACAGG...CTTGCCACACTCAACGCAATCTTAAGCAATG 2351
1371 ThrLeuSerGlyAspIleGlnThrAlaLysGlyAlaSerGlnAlaAsnAs 1387
2352 CGATACACGTTATACA..... 2367
1387 nSerAlaGlyTyrValAlaAspAlaAspGlyAsnLysIleValIleTyrAsp 1404
2367 2367
1404 eThrAspAsnLysTyrTyrGlnAlaLysAsnAspGlyThrValAspLys 1420
2368GTCAGCCCAACAGCCACCCAAA 2390
1421 ThrLysGluValAlaLysAspLysLeuValAlaGlnAlaGlnThrProAs 1437
2391 CGGCAACTTACCTTCGTCGGGCAATGCCCAAGCAACATTTAATCAAGCA 2440
1437 pGlyThrLeuAlaGlnMet...AsnValLysSerValIleAsnLysGluG 1453
2441 CATTAACGGCAACACATCGGCTTCGGCAATGCTTCAATTAATCTAAGC 2490
1453 lValAsnAspAlaAsnLysLysGlnGly..... 1462
2491 GACCAACCGCTGACAAACGCGACGCTGACGCTTCGGCAACGCTAAGC 2540
1463IleAsnGluAspAsnAlaPheValLysGlyLeuGluLysAl 1476
2541 AAACGTAAACCATTCGCACTCAACGTAATGTCCTTACCGCAATAG 2590
1476 alAspSerAspAsnLysThrLysAsnAlaValThrValGlyLysPheAla 1493
2591 CAGTATTCATTTGAAGACCGCTTACCGGACA..... 2628
1493 snAlaValAlaGlnThrProLeuThrPheAlaGlyAspThrGlyThrThr 1509
2629ATCAGCGCGGCAAGGATACGCG 2651
1510 AlaLysLysLeuGlyGluThrLeuThrIleLysGlyGlnThrAspTh 1526
2652 ATTACACTTAAAGACAGCGAATGACGCTGCCCTCAGCAGCAATTAAG 2701
1526 rAsnLysLeuThrAspAsnAsnIleGlyValValAlaGlyThrAspGlyP 1543
2702 GCAATTTAAACCTTGCAAAACGCGCACCATTTACACCTCAATTCGCGCATGCG 2751
1543 heThrValLysLeuAlaLysAspLeuThrAsnLeuAsnSerVal..... 1557
2752 CACGATGCGGAGGGGCGCAACCGCAGTGCAGACAGATGCGCGCGCG 2801
1558 ...AsnAlaGlyGlyThrLysIleAspAspLysGlyValSerPheValAs 1573
2802 CGTTTCGCGCGCTTCGCGCGCTTCCTATTATTCGTTACACCGCACTT 2851

```

1573 pserSergIyGlnAlaIalysAlasn.....ThrProValLeuS 1586
      ||| .....: |||||
2852 CGGTGAATACCCGGTTTCACACAGCTGACGGTAAAGCGCAATG.... 2895
      ||| ||| .....: |||||
1586 erAla.....AsnGlyLeuAspLeuGlyGlyValIleSer 1598
      ||| ||| .....: |||||
2896 ...AACGGTCAGGAGAACATTCGGCTTTATGTGGAACCTCTTGCGTACCG 2942
      ||| ||| ||| .....: |||||
1599 AsnValIGlyLysGlyThr.....LysAs 1606
      ||| ||| .....: |||||
2943 CAGCGCAACAATGAAGCTGGCGGAAGTTCCGAGAGCATTACACCTTGG 2992
      ||| ||| .....: |||||
1606 pThrAspAlaAlaAsnValIGlnGlnLeuAsnGlnValArgAsnLeuG 1623
      ||| ||| .....: |||||
2993 CGGTCAACATACGGGCAACGAACTGCAAGCTCGAACATATGACGTA 3042
      ||| ||| .....: |||||
1623 LysLeuGlyAsnAlaGlyAsnAspAlaAlaSpGlyAsnGlnValAsnIle 1639
      ||| ||| .....: |||||
3043 GTGAAGAGAAAGACAAACACCGCTGCGAAACCTTAATTCACCT 3092
      ||| ||| .....: |||||
1640 AlaAspIleLysLysAspProAsnSerGlySerSerSerArgThrVa 1656
      ||| ||| .....: |||||
3093 G.....CAAAACGACACGTG 3109
      ||| ||| .....: |||||
1656 LlleLysAlaGlyThrValLeuGlyLysGlyAsnAsnAspThrGlnL 1673
      ||| ||| .....: |||||
3110 ATGCGGCGCGGTGCGTACCAATC...ATCCGCAAGAGCGGCGCAT 3156
      ||| ||| .....: |||||
1673 ySLeuAlaThrGlyGlyIleGlnValIGlyValAspLysAspLysAla 1689
      ||| ||| .....: |||||
3157 CGC.....CTGCATATATCG...GTCAAGACAAAGAGCTTTCGACAA 3197
      ||| ||| .....: |||||
1690 AsnGlyAspLeuSerAsnValTrpValIleThrGlnLysAspGlySerLy 1706
      ||| ||| .....: |||||
3198 ACTGGCAAGCAGAAAGCCAAACAGCGGAAAAACAAAC..... 3240
      ||| ||| .....: |||||
1706 yLysAlaLeuLeuAlaThrTyrAsnAlaAlaGlyGlnThrAsnGlyLeuT 1723
      ||| ||| .....: |||||
3241 .....GCCAAGACCTTGCGCGCTGATGCGCGCGGCGCGCAT 3279
      ||| ||| .....: |||||
1723 hrAsnAsnProAlaGlnAlaIleAspArgIleAsnGlnGlnIleArg 1739
      ||| ||| .....: |||||
3280 GCCGTGAAAGACAGAAAGCGTTCGCAACCGCGCGGAGCGAGCGCG 3329
      ||| ||| .....: |||||
1740 PhePheHisValAsnAspGlyAsnGlnGlnProValValGln..... 1754
      ||| ||| .....: |||||
3330 GGAAGAATGTCCGCAATTATGCAGCGGAGAGAGAAAAACGGGTG.... 3375
      ||| ||| .....: |||||
1754 yArgAsn...GlyIleAspSerSerAlaSerGlyLysHisSerValAlaI 1770
      ||| ||| .....: |||||
3376 .....CAGCGCGATTAAGACACCGCTTGCGAAGAACAGCGC 3411
      ||| ||| .....: |||||
1770 leGlyPheGlnAlaLysAlaAspGlyGlnAlaAlaValAlaIleGlyArg 1786
      ||| ||| .....: |||||
3412 GAAGCGGAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGC 3461
      ||| ||| .....: |||||
1787 GlnThrGlnAlaGlyAsnGlnSerIleAlaIleGlyAspAsnAlaGlnAl 1803
      ||| ||| .....: |||||
3462 CCGCGCGGATTTGCCGCAACTGCACCAACCGACGCCCAACCGCACGC 3511
      ||| ||| .....: |||||
1803 aThrGlyAsp.....GlnS 1808
      ||| ||| .....: |||||
3512 GCGACCTGATGACCGTTAT.....GCCAATAGCGGTTTG 3546
      ||| ||| .....: |||||
1808 erIleAlaIleGlyArgThrAsnValAlaIaGlyLysHisSerGlyAla 1824
      ||| ||| .....: |||||
3547 AGTGAATTTTCCGCGACGCTCAACAGCTTTCGCGCTACAGAGAG...A 3593
      ||| ||| .....: |||||
1825 IleGlyAspProSerThrValIleAlaAspAsnSerTyrSerValGlyAl 1841
      ||| ||| .....: |||||
3594 ATTAAACCGCGTATTTGCCGAAGACCGCGCGCTTTGGAGAACGC 3643
      ||| ||| .....: |||||

```

```

1841 snAsnAsnGlnPheThrAspAlaThrGlnThrAspValPheGlyValGly 1857
      ||| ||| .....: |||||
3644 GCATCCGGGACACCAACACTACCGCTTCGACAGATTTCCGGCGCTACCG 3693
      ||| ||| .....: |||||
1858 .....AsnAsnIleThrValThrGlnSerAsnSerValAlaIle 1870
      ||| ||| .....: |||||
3694 CAACAAACCGCACTCGCGCAATCGGATGCAGAAACCTCGGACGCGG 3743
      ||| ||| .....: |||||
1870 uGly.....SerAsnSerAlaIleSerAlaGlyThrHisAlaG 1883
      ||| ||| .....: |||||
3744 GCGCGTGGGATCCGTGTTTCGACAAACCGAGCGAAGAACCTTCGAGC 3793
      ||| ||| .....: |||||
1883 LysThrGlnAlaLys.....LysSerAspGlyThrAlaGlyThrThrThr 1897
      ||| ||| .....: |||||
3794 ACGGCATCGGCACACTCGGACAGCGCTTCCACAGCGCGCTTTCGGGCA 3843
      ||| ||| .....: |||||
1898 ThrAlaGlyAlaThrGly.....ThrValLysGlyPheAlaGlyG 1911
      ||| ||| .....: |||||
3844 TACGCA.....TCGACAGGTCTACATCGCATCGCGCGCGG 3881
      ||| ||| .....: |||||
1911 nThrAlaValAlaGlyAlaValSerValGlyAlaSerGlyAlaGlnArgAl 1928
      ||| ||| .....: |||||
3882 CGCGGTTTTCAGCAGCGCGACCTTTCAGACGGCATCG 3919
      ||| ||| .....: |||||
1928 leGlnAsnValAlaAlaGlyGlnValSerAlaThrSer 1940
      ||| ||| .....: |||||

seq_name: /cgn2.6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-49

seq_documentation_block:
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-49

alignment_scores:
      Quality: 232.50      Length: 1750
      Ratio: 0.305      Gaps: 87
      Percent Similarity: 43.600      Percent Identity: 19.143

alignment_block:
US-09-303-518D-649 x US-09-268-347-49 ..
Align seg 1/1 to: US-09-268-347-49 from: 1 to: 2314

541 GATGGCGGGAATATATATCATCAAAATTAAT.....TACCCAGCGGTG 584
      ||||| ||||| .....: |||||
460 AspGlyLeuLysPheThrAspAsnSerAsnThrAlaLeuGlnSpThr 476
      ||||| ||||| .....: |||||
585 TCGT.....ATTGGGCGAGCGAGCGACATATTTGGCGATCTG 619
      ||||| ||||| .....: |||||
476 rArgIleThrLysAspLysIleGlyPheSerAsnLysAlaGlyThrVala 493
      ||||| ||||| .....: |||||
620 ATGAAATGAGCCCAATACCGGAAAGTTCAATATCATATTCGAAGTGGC 669
      ||||| ||||| .....: |||||
493 spGlnAsnLysProTyrLeuAspLysAspLysLeuLysValGlyAsn... 508
      ||||| ||||| .....: |||||
670 TATTCGTGGCTGCTGGTGGCAATACCTTGCACAAATGAGATCAGGTG 719
      ||||| ||||| .....: |||||
509 ...SerThrLeuAsnGlnGlyGlyLeuThrValAsnAsnThrIleGly 524
      ||||| ||||| .....: |||||
720 TGGC..... 723

```


Ratio:	0.693	Gaps:	35
Percent Similarity:	41.032	Percent Identity:	25.061

```
alignment_block:
US-09-303-518D-649 x US-09-413-814-78 .
```

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

[illegible][illegible]

```

519 AAlaIaProaArgIyAlaAlaAspAspValGlnArgInglYArgAl 555
3818 TTGGCCAGCGCGCGTTTGGGCA.....ATACGGC 3849
      ||| ||||| |||||
535 aValProglYArgArgAlaArgglYProglYAlaValProAspValArgP 552
3850 ATCGACAGGTTTACATCGGCATCGCGCGCGGGGTTTACGACGG 3899
      ||| ||| ||| |||||
552 roArgAlaProaArgIYArgglYProaRg...ArgglYglYLeuGlYAla 567
3900 CAGGCTTTCAGCGCATCGGAGCAAAATCCCGCGCGGCTGTGAT 3949
      :||| :||| :||| :||| :||| :||| :||| :|||
568 ArgProglYglYglYProaArgAlaValArgProglYProaArgAlaLar 564
3950 ACGCATTCAGGC..... 3962
      :||| :||| :||| :||| :||| :||| :||| :|||
584 oProaArgAlaGlYhIsProglYAlaProaArgAlaArgHIsArgAla 601
3963 ..ACGATACCGCGCGGTTTGGCGGATTCGCAATCGAACCGCATCGG 4010
      ||||| ||||| ||||| |||||
601 rArgAlaProaArgValArgArgLeuValGlYArgArgLeuArgArg 617
4011 .....CGCAACGGCTATTTCG.....CCAA 4033
      ||||| ||||| ||||| |||||
618 ArgAlaArgArgAlaLeuArgArgLeuArgAlaGlYProAlaPheProAl 634
4034 AAGCGATTCACCGCTACGAAACGTCATATCGCCACCGCGCTTGA 4083
      :||| :||| :||| :||| :||| :||| :||| :|||
634 aAlaGlYAlaProglYAlaValaArgArgLeuArg...SerProAlaG 650
4084 TTCACCGCTACCGCGCGGCTTACGCAAGATTATTCATCAACCGCG 4133
      :||| :||| :||| :||| :||| :||| :||| :|||
650 lYValAlaValaArgArgIYProglYglYArg..... 660
4134 GCACACATTTTCATCAAGCGCTTTTACGCTTCTCTATACGATCGCG 4183
      ||||| ||||| ||||| |||||
661 .....AlaProValLeuAspAspVal 668
4184 CTTCGGCAAGTCCGAAACAGCGCTCAATACCGCGTATTGGCTAGAT 4233
      :||| :||| :||| :||| :||| :||| :||| :|||
668 aLeuGlYArgAlaAlaProglYProaArgAlaGlYArgProaArgAlaGl 665
4234 TTCGGCAAAACCGCGAGTCCGGAATGGCGCTAAACCGCAATCAAG 4283
      :||| :||| :||| :||| :||| :||| :||| :|||
685 lYAlaLysAspValAlaArgglYAlaAlaGlYProaArgArgglYArg 701
4284 TTTCACGCTGC..... 4295
      |||
702 AlahIsProaAspArgInglYValLeuArgAlaGlYlYAspLeuProle 718
4296 .....CTTCACGCTGCCGCGCAAGCGCGCA.....ACTG 4329
      ||| ||| :||| :||| :||| :||| :||| :|||
718 uHIsGlYProAlaArgArgValaGlInAlaArgProAlaProAlaHIsArg 755
4330 GAAGCGCAACACAGCGCGGCTCAATTAAGCTACCG 4367
      :||| :||| :||| :||| :||| :||| :||| :|||
735 laArgglYProaArgArgglYhIsArgArgArgInPro 747
seq_name: /cgn2_6/prodata/1/laa/6B_COMB.pep:US-09-377-155-33

```

```

; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

alignment_scores:
  Quality: 229.50      Length: 1495
  Ratio: 0.339         Gaps: 72
  Percent Similarity: 45.284      Percent Identity: 19.130

alignment_block:
US-09-303-518D-649 x US-09-377-155-33 ..
Align seg 1/1 to: US-09-377-155-33 from: 1 to: 2353

172 GCCGAAATTAAGCAAGTTTGCAGTCGCGGCGAAGATATTAAGATT 221
||| :||| :||| :||| :||| :||| :||| :|||
436 AlaAspThrAspGlYAlaLeuInglYlIleSerlYAspGlInglVal.. 451
222 CAACAAAAAGGGAGTTGTC.....GCCAATCAATG...A 256
||| ||||| ||||| ||||| :||| :||| :||| :|||
452 ...LysAlaGlYglYThrValThrPheLysAlaGlYlYsAsnLeuLysV 467
257 CAAAAGCCCCGATGATTTTCTGTGTCGCCGTAACGGCGTGGCG 306
:|||| :||| :||| :||| :||| :||| :||| :|||
467 allYsGlInAspGlYAlaAsnPheThrYrSerLeuGlInAspAlaLeuThr 483
307 GCATTGTGTGCG..... 318
:|||| :||| :||| :||| :||| :||| :||| :|||
484 GlYLeuThrSerlIeThrLeuGlYlYThrThrAsnGlYlYAsnAspAl 500
319 .....GATCAATATATTGTGAGCGTGCACTAATACG 349
||| :||| :||| :||| :||| :||| :||| :|||
500 alYsThrValIleAsnLysAspGlYLeuThrIleThrProAlaGlYAsnG 517
350 GCGGCTATACAACTGTGATTGTTGGCGGAAGAAATCCCATCA 399
||| :||| :||| :||| :||| :||| :||| :|||
517 lY.....GlYThrThrGlYThrAsn..... 523
400 CATCGTTTACTTATAAATGTGAACGGAATATTAAGCAGGAGC 449
||| :||| :||| :||| :||| :||| :||| :|||
524 .....ThrIleSerValThrLys...AspGlYlYlYsAlaGlYAs 536
450 TAAAGCCATCCTTATGGCGGCAATATCAATATGCGGCTTGCAATAAT 499
:|||| :||| :||| :||| :||| :||| :||| :|||
536 nLysAla..... 538
500 TTGTACAGATGCAGAACCTGTGAATGACCACTTATATGATGGCGG 549
:|||| :||| :||| :||| :||| :||| :||| :|||
539 .....lIeThrAsnValAlaSerlYLeu 546
550 AATATATGATCAAAATTAATTAACCGTACCGTGTCTATTGGGCGAG 599
:||| :||| :||| :||| :||| :||| :||| :|||
547 ArgAlaIYrAspAspAlaAsnPhe..... 554
600 CAGGCAATATTGGCGATGTGATGAAGATGAGCCCAATAC.....CGCG 643
||| ||| ||||| :||| :||| :||| :|||
555 .....AspValLeuAsnAsnSerAlaThr 563
644 AAAGTTCAATATATTTGCAAGTGGCTATTTCTGGCTGTGGTGCAAT 693
:||| :||| :||| :||| :||| :||| :||| :|||
563 sPLeuAsnArgHIsValaGlYAspAlaIYrLys..... 573
694 ACCTTTGCACAAAATGATCAAGTGTGTCAGACAGTCACTAGTAGTGA 743
||| :||| :||| :||| :||| :||| :||| :|||
574 .....GlYLeuLeuAsnLeuAsnGlYlY 581

```


1089 u.....ValasnaIagIuGIyleuAlaThralaLeuAsna 1101
2336 ATCTTAGT.....GCAATGGCGATACGCT 2361
1101 snLeuSerTTPThralaIalysAlaAspIysTyrAlaAspGIyIuSerGIu 1117
2362 TATACAGTACAGCCAAACGCCACCCAAACGCAACCTTACGCTC 2406
1118 GIyIuThraspGIuIuValIyAlaGIyAspIysValIThrPheIysAl 1134
2407GTGGCAATGCCCAAGCACATTTAAT..... 2433
1134 aGIyIyAsnLeuIyValIyGInserGIuIyAspPheThryIserI 1151
2434 ..CAGGCACATTAAGCGCACACATCGGCTTG.....GGCAATGCT 2475
1151 euGIAspThIleuthrGIyleuThrSerIleThrIleuGIyGIyThAla 1167
2476 TCATTTATCTAAGCAGCACGCCGTACAAACGCG.....ACTCT 2516
1168 AsnGIyAspAsnAspThrGIyThrValIleAsnIyAspGIyleuThrIL 1184
2517 GACGCTTTCGGCAGACGCTAAGCAACGTAAGCATTCCGACTACAG 2566
1184 eThrIleuAlaAsnGIyAlaAlaIaGIyThr....AspAlaSerAsnG 1199
2567 GTAAAT.....GTCTCCTAAGCGGATTAAGCA 2592
1199 IyAsnThrIleSerValIThrIyAspGIyIleSerAlaGIyAsnGIu 1215
2593 GTATTCATTTTGAAGCAGCGGCTTACGCGACAAATCAGCGGCGCA 2642
1216 IleThrasnValIySerAlaIleuIyThr....TyrIyAspThrGI 1230
2643 GGATACGCGCATTAACCTTAAAGCAGCAATG.....ACGTCGCGCT 2686
1230 nasThralaAspGIuThrGIAspGIuThrPheIleAlaValIyAs 1247
2687 CAGCAGCAGGAATTAAGCAATTAACCTTGACAGCGCCACCATTA 2733
1247 snAlaAsnGIuValGIuPheValGIyIyAsnGIyAlaThrValSerAla 1263
2734 ...CTCAATTCGCGCTATCGCCAC.....GATGCGGACGAGGCG 2768
1264 IySthraspAsnAsnGIyIyShISthValIThrIleAspValAlaGIuAl 1280
2769 GCAACGCGGCGAGTGG.....ACAGATCG.....CCGCGCG 2800
1280 aIyValaIyAspGIyIeugIuIyAspThraspGIyIySILeIySLeuI 1297
2801 GCCGTTGCGCGGCTTCCGCGCTTCCATTAATCCGTT.....ACACG 2844
1297 ySValAspAsnThraspGIyAsnAsnLeuThrValaAspAlaThrIyS 1313
2845 CCAACTTCGGTA...GAATCCGTTTCAACAGCTGACG..... 2880
1314 GIyIaSerValaIaIyGIyGIuPheAsnAlaValIThrThraspAlaTh 1330
2881GTAACGGCAATTAAGCGGTGAGGAACATTCGCT 2917
1330 rThAlaIaGIyThrasnAlaAsnGIuIaGIyIyValIaValIySG 1347
2918 TTATGCGGAACCTTCGCTACCGCAGCAGCAAAATGAAGTCGCGAA 2967
1347 IySerAsnGIyAlaThralaThrGIuThraspIySlySlyValAla... 1362
2968 AGTTCGGAAGCACTTACACCTTGGCGGTCAACATACGCGCAACGACC 3017
1363 ...ThrValGIyAspValaIaIyAlaIleAsnAspAlaIaThr.... 1376
3018 TGCAAGCCTGACACATTAAGCGGTAGTGAAGAAACACACAAACGCG 3067
1377PheValIyValGIuAsnAspSerAlaThrI 1388

3068 TGTCGAAACCTTAATTACCTGCAAAAGCAGACGTCGATCCGCG 3117
1388 IeAspAspSerProThraspAspGIyAlaAsnAspAlaLeuIyAlaGIy 1404
3118CGGTGGCGTTTACCACTCATCCGCAAGACG 3149
1405 AspThrIleuThrLeuIyAlaGIyIyAsnLeuIySlyArgAspGI 1421
3150 CGAG.....TTCGCGCTCATTAAT.....CCGTCGAAAGACAG 3184
1421 IyAsnIleThrPheAlaLeuAlaAsnAspLeuSerValIySserAlaT 1438
3185 ACCTTTCGCAAACTC.....GGCAAGCAGAGAACCC 3216
1438 hrValaSerAspIySleuSerIeugIyThrasnGIyAsnIyValaSnIle 1454
3217 AAAAAACAGCGGAA.....AAGACAAACGCGCAAGCCT 3251
1455 ThrSerAspThrIyGIyIeAsnPheAlaIyAspSerIySthGIyAs 1471
3252 TGACGCGCTGATTCGCGCGCGGCGATCCGTCGAAAGACAGAACG 3301
1471 pasPalaAsnIleHISleuAsnGIyIleIaSerThrIleuThraspThI 1488
3302 TT.....GCCGAACCGCGCGCAGCAGCGCGGCAAAATGTCGCAAT 3345
1488 euleuAsnSerGIyAlaThrThrasnLeuGIyIyAsnGIyIle..... 1502
3346 ATGACAGCGGAGGAAAGAAAAACGGGTGACGCGGATTAAGACACCGC 3395
1503ThraspAsnGIuIySlySlyArgAlaIaIaSerValIyAspValle 1517
3396 CTTCGGGAACAGCGGGAAGCGAAACCGCGCGCTTCCACCGGCTTCC 3445
1517 uAsnAlaGIyTTPAsnValaIaGIyValIySPTAlaSer..... 1530
3446 CCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACTGCAACCCACG 3495
1530 1530
3496 CAGCCCCAACCCGACGCGACCTGATCAGCCGTTATGCCAATAGCGGTT 3545
1531AlaAsnAsnGIuAla 1535
3546 GAGT.....GAATTTCCGCCACGCTCAACAGCGTTTCCGCGTACAG 3589
1535 IGIuAsnIleAspPheValaIaThrTyrAspThrValaAspPheValSerG 1552
3590 ACGAATTGAC.....CGCGTATTTGCCGAAGACCGCGCAC..... 3627
1552 IyAspIyAspThrThraserValIThrValGIuSerIyAspAsnGIyS 1568
3628GCCGTTTGACCAAGCGCATCCGCAACACCA 3659
1569 ArgThrGIuValIySleGIyAlaIySthraserValIleIySAsp.... 1583
3660 ACACCTACCTTCGCAAGATTTCGCGCTTACCGCCCAACAAACGACCTGC 3709
1584 ...HisAsnGIyIySleuPheThrGIy...IySleuIeugIyAspAla 1598
3710 GCCAAATCGTATCGAG.....AAAACCTCGGACG 3741
1598 snAsnAsnGIyAlaThrValaIThrGIuThraspGIyIyAspIeugIySn 1614
3742 GGGCGCGTTC.....GGCATCTGTTTTCGCAACCGG..... 3774
1615 GIyIeugIyThralaIyAlaValIleAspAlaValaIySlyAlaGIyTr 1631
3774 3774
1631 parGIyIySthThrThrGIyAlaAsnGIyIuAsnAspAspPheAlaThrV 1648

985 AATACCTCTTTTAAAGAGATATATGACAGGAAAAATCAATGACCA 1034
 |||:|||||
 644 LysHisThr..... 646
 1035 ACATGACACAAATCTCTGCCCTAATAGATTAAAAACGAAACCGTTCAAT 1084
 646 646
 1085 TGTATTAATGTTCTTTATCCGACAGACGAAACCTTTATCATGCT 1134
 |||:|||||:|||||
 647 ..IleThrValSerValIleGluThrLysAlaAspCysGlyLeuGluLys 662
 1135 GCAGGTGCTGTCACACTTATCGACCCAGACTCAATTAATGAGAAAAATAT 1184
 |||:|||||:|||||
 663 AspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnVal 679
 1185 TTCCTTTAT..... 1194
 679 IleThrValGlyAsnAsnGlyThrAlaValThrLysGlyLysPheGluT 696
 1195GACGAGGAAAAAGCGCAATGTACTT... 1221
 696 hrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLys 712
 1222ACCAGCAAC..... 1230
 713 AspAlaThrAlaAsnAspAlaAspLysValAlaThrValLysAspVal 729
 1231ATCAATCAAGTGTGTCGAGATTTATTTCCAGAGAGACTT 1270
 |||:|||||:|||||
 729 LAlaThrAlaIleAsnSerAlaIleAlaThrPheValLysThrGluAsnLeuT 746
 1271 TTACGCTCTCCCTGAAATAACGAACTGTGCAAGCGCGCGCTTCAT 1320
 |||:|||||
 746 hrThrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAla 762
 1321 ATCAGTGAAGACAGTACCTTACTTGAAAGTAAACGCGCTGCAACGA 1370
 |||:|||||
 763 LeuLysAlaGlyAspThrLeuThrPheLysAlaGly.....LysAsnLe 777
 1371 CCGCTGTCCAAATCGCAAGGC..... 1395
 |||:|||||
 777 uLysValLysAlaArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnL 794
 1396ACGCTGCACGTTCAAGCCAAA 1416
 |||:|||||
 794 euGluValLysThrAlaLysValSerAspThrLeuThrIleGlyLysAsn 810
 1417 GGGGAAACCAAGGCTG.....ATCAGCGT 1442
 |||:|||||
 811 ThrProThrGlyGlyThrThrAlaThrProLysValAsnIleThrSerTh 827
 1443 GGGGACGCTACAGTCATTTTGATGACAGGACGATTAAGGCAAAA 1492
 |||:|||||
 827 rAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGlySerL 844
 1493 AACCAAGCCTTT.....AGTGAATC 1512
 |||:|||||
 844 yAsnValIleThrLeuLysGlyIleAlaThrThrLeuThrGluProSerAla 860
 1513 GGCTTGGTCACGGGAGGAGTGCATGCACTGATGCCGATATATCGT 1562
 |||:|||||
 861 GlIleLysSerSerHisValAspLeuAsnValAspAlaThrLysLysSe 877
 1563 CAACCCCGCAACTATTTTCGCTTCGCGGCGGA...CGTTGGATT 1609
 |||:|||||
 877 rAsnAlaIleSerIleGluAspValLeuArgAlaGlyTyrAsnIleGln 894
 1610 TAAACGGGATTCGCTTTCGTCACCGATTCATAAATACCGATGAAGG 1659
 |||:|||||
 894 LysAsnGlyAsnAsnValAspTyrValAlaThrThrLysAspThr..... 907

1660 GCGATGATGTCTCAACCAATCAAGCAAAAGATCCACCGTTACATAC 1709
 |||:|||||:|||||
 908ValAsnPheThrAspAspSerThrGlyThrThrValTh 921
 1710 AGGCATTAAGATATTGCTACAAACGGCAATTAACAACGCTTGATAGA 1759
 |||:|||||
 921 rValThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAla 938
 1760 AAAAAGAAATGCC.....TACACGCT...TGGTTGGCCAGAAAGAT 1800
 |||:|||||
 938 yThrSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAsp 954
 1801 ACGACCAAAACGACGCGGCTCAACCTGTTTACACCGCCGCGCAGA 1850
 |||:|||||
 955 LeuLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAs 971
 1851 AGACCGCACCCCTGCTCTTCGCGCGAACA.....AATT 1885
 |||:|||||
 971 pThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaValAsnL 988
 1886 TAAACGGCAACATCACGCAAAACGCAAACTGTTTTCAGCGCAGA 1935
 |||:|||||
 988 ySerGlyTyrPargValThrGlyGluGlyAlaThrAlaGluThrGlyAla 1004
 1936 CCAACACCGCAGCGCTACATCATTTAAACGACCATTTGGTGCAAAAGA 1985
 |||:|||||
 1005 ThrAlaValAsnAlaGlyAsn.....AlaGlu 1013
 1986 GGGCATTCCTCGCGGGAATGCTGTGCGGCAACGACTGATCAACCGCA 2035
 |||:|||||
 1013 uThrValThrSerGlyThrSerVal..... 1021
 2036 CATTTAAACGGCAAACTTCCAAATTAAGGCGGACAGCGGTGTTCC 2085
 |||:|||||
 1022AspPhe.....LysAsnGlyAsnAlaThrThrAla 1031
 2086 CGCAATGTGCCCAAGAGTGAAGCGCATTTGGCATTTGAGCAATCACGCCA 2135
 |||:|||||
 1032 ...ThrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAs 1047
 2136 AGCAGTTTGTGTGTCGACCGCATCAACGCAACAAATCTGTACAGCT 2185
 |||:|||||
 1047 nValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal.....A 1062
 2186 CGGACTGACGCGGCTGTGACAAATGTGTGCAAAAACCATTAACGAGCAT 2235
 |||:|||||
 1062 LAspThrThrThrLeu.....ThrValThrGlyGly 1072
 2236 AAAGTATGTCTCATTTGACTAAGACCGCATCAGCGGCAATGTGATCT 2285
 |||:|||||
 1073 LysValSerValProAlaGlyAlaAsnSerValAsnAsnLysLysLe 1089
 2286 TGGCGATCACGCTCATTTTAATCTCACAGGCGTCTCCACACATCAAGGCA 2335
 |||:|||||
 1089 u.....ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1101
 2336 ATCTTAGT.....GCAAAATGGCGATACAGCT 2361
 |||:|||||
 1101 snLeuSerThrThrAlaLysAlaAspLysTyrAlaAspGlyLysSerGlu 1117
 2362 TATACAGTACGACACACAGCCCAACCAAGCGCAACTTAGCTC..... 2406
 |||:|||||
 1118 GlGlyLysThrAspGlnGluValLysAlaGlyLysValLysValThrPheLysAl 1134
 2407GTGGGCAATGCCCAAGCAACATTTAAT..... 2433
 1134 agLysAsnLeuLysValLysGlnSerGluLysAspPheThrThrSerL 1151
 2434 ..CAAGCCATTAACGCAACACATCGGCTTCG.....GGCAATGCT 2475
 |||:|||||
 1151 euGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAla 1167
 2476 TCATTTAATCTAAGCAGCACGCGCTACAAAACGCG.....AGTCT 2516


```
1168 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIle 1184
1169 GAGCTTTCGGCAGCGTAAGCAAAAGCAAGCAATCCGACCTCAACG 2566
1170 eTThrLeuAlaAsnGlyAlaIleAlaGlyThr.....AspIleSerAsnG 1199
2567 GAAAT.....GTCCTCCATCCGATAGGCA 2592
1199 LysThrIleSerValThrLysAspGlyIleSerAlaGlyAsnGly 1215
2593 GAATTCATTTGAAAGCAGCGCTTACCGCAAAATCAGCGGCGCA 2642
1216 IleThrAsnValLysSerAlaLeuLysThr.....TyrLysAspThrGly 1230
2643 GAGTACGGCATTAACCTAAAGACAGCAATG.....ACGCTGCCGT 2686
1230 nsnThrAlaAspGluThrGlnAspLysGluPheHisAlaAlaValLysA 1247
2687 CAGGCAGGAATTAGGCAATTAACTTAACTGACACAGCCACATTAACA... 2733
1247 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1263
2734 ...CTCAATTCGCGCTATGCGCAC.....GATGCGCAGAGGCG 2768
1264 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluVal 1280
2769 GCAACCGCGCAGTCCG.....ACAGATGCG.....CCGCGCG 2800
1280 aLysValGlyAspGlyLeuGlyLysAspThrAspGlyLysIleLysLeu 1297
2801 GCGGTTCGCGCGCTGCGCGCTTCCCTATATCCGT.....ACAGCG 2844
1297 yValAspAsnThrAspGlyLysAsnLeuLeuThrValAspAlaThrLys 1313
2845 CCAACTTCGTA...GAATCCGTTTCAACAGCGCTGACG..... 2880
1314 GlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaThr 1330
2881 .....GTAAAGCGCAATTGAAGGTCAGGAGACCTCCGCT 2917
1330 rThrAlaGlnGlyThrAsnAlaAsnGluThrGlyLysValValValLys 1347
2918 TTTATGTCGAACCTTCGCGCTACCGAGCAGCAAAATTGAAGCTGGCGGA 2967
1347 LysSerAsnGlyAlaThrAlaThrGluThrAspLysLysValAla... 1362
2968 AGTTCCGAGGCACTTACACCTTGCGGCTACACATACCGCAGACGACC 3017
1363 ...ThValGlyAspValAlaLysAlaIleAsnAspAlaIleThr..... 1376
3018 TGCAGCGCTGACAATGTACGCTAGTGAAGAAAGCAACAAACCGG 3067
1377 .....PheValLysValGluAsnAspAspSerAlaThrI 1388
3068 TGTCGGAACCTTAATTCACCTCGCAAAAGCAACAGCTGATGCGCGC 3117
1388 leAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAlaGly 1404
3118 .....GCGTGGCGTTACCACTCATCCGCAAAACGCG 3149
1405 AspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGly 1421
3150 CGAG.....TTCGCGCTGCATAT.....CCGCTCAAGAACAAAG 3184
1421 yLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSerAla 1438
3185 AGCTTTCGACCAATC.....GGCAGGACGAGAGCC 3216
1438 hrValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIle 1454
3217 AAAAGACAGCGCGAA.....AAAGACAGCGCGCAAAAGCT 3251
```

```
1455 ThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThrLys 1471
3252 TGACGCGCTGATTCGCGCGCGCGCGATGCGCGTAAGCAAAAGCAAGCG 3301
1471 pAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThr 1488
3302 TT.....GCCGACCGCGCGCGCGCGCGCGGGAATGTCGGCAT 3345
1488 euleAsnSerGlyAlaThrThrAsnLeuGlyLysAsnGlyLe..... 1502
3346 ATGCAAGCGGAGAGAGAAAAACGGCTGACGCGGATTAAGACACCGC 3395
1503 .....ThrAspAsnGlyLysArgAlaAlaSerValLysAspVal 1517
3396 CTGGCGGAACAGCGCGAGGAAACCGCGCGCTACCAACCGCCTTC 3445
1517 uAsnAlaGlyTyrAsnValArgGlyValLysProAlaSer..... 1530
3446 CCGCGCGCGCGCGCGCGCGGATTTGCCGCAACTGCAACCCCAACG 3495
1530 ..... 1530
3496 CAGCCCCACGCGACGCGACCTATCAGCGGTATGCCAATAGCGTT 3545
1531 .....AlaAsnAsnGluVal 1535
3546 GAGT.....GAATTTCCGCGCAGCTCAACAGCGTTTTCGCGCTACAG 3589
1535 GluAsnIleAspPheValAlaThrTyrAspThrValAspPheValSerG 1552
3590 ACGAATTAGAC.....CGCGTATTTGCCGAGACCGCGCGCAC..... 3627
1552 LysAspLysAspThrThrSerValThrValGluSerLysAspAsnGlyLys 1568
3628 .....GCCGTTGGACAGCGGCAATCCGCGCACACCA 3659
1569 ArgThrGluValLysIleGlyAlaLysThrSerValIleLysAsp..... 1583
1584 ...HisAsnGlyLysLeuPheThrGly...LysGluLeuLysAspAla 1598
3710 GCCAATCGGTATGCAG.....AAAACCTGGCAGC 3741
1598 snAsAsnGlyValThrValThrGluThrAspGlyLysAspGluGlyAsn 1614
3742 GGGCGCGTC.....GGCATCCTGTTTGGCAGACACGCG..... 3774
1615 GlyLeuValThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTr 1631
3774 ..... 3774
1631 pArgValLysThrThrGlyAlaAsnGlyGluAsnAspPheAlaThrV 1648
3775 .....ACGGAACACCTTCGACGACGCGCATGCGCACTCGCA 3813
1648 aLAlaSerGlyThrAsnValThrPheAlaAspGlyAsnGlyThrThrAla 1664
3814 GCGCTTGC.....CACGCGCGCTTTTGGGCAATACGCGATGA 3854
1665 GluValThrLysAlaAsnAspGlySerIleThrValLysTyrAsnVal.. 1680
3855 CAGGTTTACATGCGCATGCGCGCGCGGCTTTAGACGCGCGACCC 3904
1681 .....LysV 1682
3905 TTTCAGACGCGCATC.....GAGGCAAAATC 3930
1682 aLAlaAspGlyLeuLysLeuAspGlyAspLysIle 1693
seq_name: /cgn2_6/ptodata/1/iaa/6b_COMB_pep:us-09-669-974-33
seq_documentation_block:
```


1371	CGCGCCGTCGCAAAATCGGCAAGGC.....	1395
777	uysvalysatgaspgllylsasnllethrPhepspleuAlalsasnl	794
1396ACGTCGACGTTCAAGCCAAA	1416
794	eucIuVallystrhrAlalsValserAspThrleuThrIleclgLYasn	810
1417	GGGGAAAACCAAGCTCG.....ATCAGCGT	1442
811	thrProthrgLYcLYthrThrAlaThrProlysValsnllethrSerth	827
1443	GGGGCAGCGGTACAGCATTTTGGATGCACAGCAGACATGAAGCAAA	1492
827	rAlaspgLYleuAsnPhenAlalsgLYthrAlaAspAlaserlyserL	844
1493	AACACGCTTT.....AGTGAATC	1512
844	ysasnVallyrLeuysgLYleAlaThrThrleuThrgLProserAla	860
1513	GGCTTGTCAGGGGAGGGGCTAGCGTGCAGTGAATGCCATATCGATT	1562
861	gLYAlaLysSerSerthlsValAspleuAsnValAspAlaThrlyLysSe	877
1563	CAACCCGCAACACTATTTGGCTTTGGCGGCGGA...CGTTGGATTT	1609
877	rAsnAlaAlaserIleclgAspValleuThrgrAlagLYrpsnIleclg	894
1610	TAAACGGCGATTCGCTTTCGTTCCACCGTATTTCAAAATACCGATGAAGG	1659
894	LYasnGLysAsnValAspTYrValAlaThrTYrAspThr.....	907
1660	GGCATGATTGTCAACACATGACAACAAGAATGCACCGTTACCATTTAC	1709
908ValAsnPhethrAspAspserThrgrLYthrThrValTh	921
1710	AGGCAATTAAGAATATTGTCTACAACCGCGCAATGAACAACGTTGATGA	1759
921	rValThrGLnLYsAlaAspGLyLysGLyAlaAspVallystrIleclgAla	938
1760	AAAAAACAATTGCC.....TCAACGCT..TGSTTTGGCGAGAAAGAT	1800
938	ysThrSerValIleLysAspPhlAsnGLyLysLeuPhethrGlyLysAsp	954
1801	ACGACCAAAACGAGAGCGGGGCTCAACCTGTGTACAGCCCGCGCGA	1850
955	LeuLYAspAlasAlasAsnGLyAlaThrValSerGIuAspAspGLyAs	971
1851	AGACCGCACCCCTGCCTTCGCGGGCGAACA.....AATT	1885
971	pThrgrLYthrGLyLeuValThrAlaLysThrValIleAspAlaValAsnL	988
1886	TAAAGGCAACATCAACGCAAAACAAAGCGAAACGTTTTTCAGCGGCGA	1935
988	ysSerlylYrpaYValThrgrLYuGLyAlaThrAlaGLuThrGLyAla	1004
1936	CCAACACCGCAGCCCTCAATCATTTTAAACGACATTGGTCGCAAAAAGA	1985
1005	ThrAlaValAsnlaGLysLn.....AlaGL	1013
1986	GGGCAATTCTCGCGGGGAATGTGTGGGACACACACATCGATCAACGGCA	2035
1013	urnValThrSerGLyThrSerVal.....	1021
2036	CATTTAAGCGGAAAACTTCCAATTTAAAGCGCGACAGCGGTGGTTTCC	2085
1022AspHe.....LYsAsnGLyAsnAlaThrThrAla	1031
2086	CGCAATGTTGCCAAAGTGAAGAAGCGCATTTGGACCAATCAACGCCCA	2135
1032	...ThrValSerLYsAspAsnGLyAsnIleAsnValLYsTYrAspValAs	1047

2136 AGCAATTTTGTGGTGGCAACGCAATCAAGCAAGCAACATCGTACAGCTT 2185
1047 nvalGlyaspGlyLeuLysIleGlyAspAspLysLysIleVal.....A 1062
2186 CGGACTGCGAGCGGGCTCGCAAAATGTGTGCAAAAAACCATTAACCGAGAT 2235
1062 laaspThrThrIleLeu.....ThValThcGly 1072
2236 AAGATGATTGGCTTCATTGACTAAGCGCAGATCAAGGGCAATGTGCATCT 2285
1073 LysValSerValProAlaGlyAlaAsnSerValAsnAsnAsnLysLysLe 1089
2286 TGGCATCAACGCTCAATTAAATCTACAGGGCTGGCCACTCAACAGGCA 2335
1089 u.....ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1101
2336 ATCTTAGT.....GCAAAATGGCGATACAGT 2361
1101 snLeuSerThrPheAlaLysAlaAspLysTyrAlaAspGlyGluSerGlu 1117
2362 TATACAGTACGCGACACAGCGCACCCCAAAAGCGACCTTACGCTC.... 2406
1118 GlyGluThrAspGluGluValLysAlaGlyAspLysValThrPheLysAl 1134
2407GTGGCAATGCCCAAGCAACATTAT..... 2433
1134 agLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1151
2434 ..CAACGCCATTAAAGCGCACACATCGGCTCG.....GGCAATGCT 2475
1151 euGlnAspThrLeuThrGlyLeuThrSerLleThrLeuGlyGlyThrAla 1167
2476 TCATTTAATCTAAAGCAGCACCGCTACAAAAGCGC.....AGTCT 2516
1168 AsnGlyArgAsnAspThrGlyThrValLleAsnLysAspGlyLeuThrIL 1184
2517 GACGCTTTCGGGCAACGCTAAGCGAAAGCTAACCCATTCCGACTCAACG 2566
1184 eThrLeuAlaAsnGlyAlaAlaAlaGlyThr....AspAlaSerAsnG 1199
2567 GTAAAT.....GTCTCCCTACCGGATTAAGCA 2592
1199 LysAsnThrIleSerValThrLysAspGlyLleSerAlaGlyAsnLysGlu 1215
2593 GATATCCATTTTAAAGCAGCGCTTACCGGCAAAATACAGCGCGGCA 2642
1216 IleThrAsnValLysSerLleLeuLysThr....TyrLysAspThrGlu 1230
2643 GGATACGCGATTACACTTAAAGACAGCGAATG.....ACGTCGCGT 2666
1230 nasnThrAlaAspGluThrGlnAspLysGluPheHisAlaIleValLysA 1247
2687 CAGGCAACGGAATTAGCAATTTAAACCTTGACACAGCCACCATTAACA... 2733
1247 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1263
2734 ..CTCAATTCGCGCTAAGCGCAC.....GATGGCGCAGCGGC 2768
1264 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1280
2769 GCAAAACCGCGAGTGC.....ACAGATGC.....CCGCGCC 2800
1280 AlLysValGlyAlaAspGlyLeuGluLysAspThrAspLysLysLysLysLeuL 1297
2801 GCCGTTGCGCGCTTGGCGCGCTTCCCTATATTCGTT.....ACACCG 2844
1297 yValAspAsnThrAspGlyAsnAsnLeuThrValAlaSpAlaThrLys 1313
2845 CCAACTTCGTA...GAATCCGTTTCAACAGCGAGAC..... 2880
1314 GlyAlaSerValAlaLysGluLysPheAsnAlaValThrThrAspAlaThr 1330
2881GTAAACGGCAAAATGAAACGCTGAGGAACATTCCGCT 2917

```

1330 rThralagInglyThrasnAlaasnGluArgglyLysValValValLysG 1347
      ::|||::: ::|||::: :::
2918 TTATGTCGAACCTTTCGGCTACCGCAGCAAAATTGACGTGGCGGAA 2967
      ::: ::: ::|||::|::|::|::|::|::|::|::|::|::|::|
1347 LysSerAsnGlyAlaThrAlaThrGluThrAspLysLysValAla... 1352
      ::: ||| ::: |||::|::|::|::|::|::|::|::|::|::|
2968 ACTCCGAAGCGCATTCACCTTGGCGGTCAACAATACCGCAAGAAC 3017
      ::: ||| ::: |||::|::|::|::|::|::|::|::|::|::|
1363 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaIaThr... 1376
      ::: ||| ::: |||::|::|::|::|::|::|::|::|::|::|
3018 TGCAGCCTCGACAAATTGACGGTGTGAAGAAAGAAACACAAACCC 3067
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1377 .....PheValLysValGluAsnAspAspSerIaThrI 1388
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3068 TGTCCGAACCTTAATTTCACCTCGCAAAAGCAACGTCGATCGCGG 3117
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1388 LeAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAlaGly 1404
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3118 .....GCGTGGCGTTACCAACTCATCCGCAAAAGCG 3149
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1405 AspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspG 1421
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3150 CGAG.....TTCCGCTGCATTAAT.....CCGCTCAAAAGACAG 3184
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1421 YLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaT 1438
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3185 AGCTTTCGACAAACTC.....GGCAAGCGACGAAGCC 3216
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1438 hrValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIle 1454
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3217 AAAAAACAGCGGGAA.....AAAGACAAAGCGCAAGCCT 3251
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1455 ThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAs 1471
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3252 TGACCGCTGATTTGGCGCGCGGATCGCTGCAAAAGCAGAAAGCG 3301
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1471 PAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrL 1488
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3302 TT.....GCGGAACCGCGCGCGAGCGCGGGGAAATGTGCGCAT 3345
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1488 euleuAsnSerGlyAlaThrThrAsnLeuGlyLysAsnGlyIle..... 1502
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3346 ATGAGGGCGGGAAGAAAGAAAGGGGTGCGAGCGGATTAAGACACCC 3395
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1503 .....ThrAspAsnGluLysLysArgAlaAlaSerValLysAspValLe 1517
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3396 CTGGCGGAACAGCGCGAAGCGGAACCGCGCTACACCGCTTCC 3445
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1517 uAsnAlaGlyTrpAsnValArgGlyValLysProAlaSer..... 1530
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3446 CCCGCGCGCGCGCGCGCGGATTTGCGCAATGCAACCCCAACG 3495
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1530 ..... 1530
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3496 CAGCCCCAAGCGAAGCGGACCTGATCAGCCCTTATGCCAATACGGTT 3545
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1531 .....AlaAsnAsnGluIva 1535
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3546 GAGT.....GAATTTCCGCGACGCTCAACAGCGTTTTCGCGCTACAG 3589
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1535 LgluAsnIleAspPheValaIaThrItyrAspThrValAspPheValSerG 1552
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3550 ACGAATTAGAC.....CGCGTATTGGCGGAAGCGCGCGCAAC... 3627
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1552 LysAspLysAspThrThrSerValThrValGluSerLysAspAsnGlyLys 1568
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3628 .....GCGTTTGACAAGCGCGCATCCGCGGACACCA 3659
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1569 ArgThrGluValLysIleGlyAlaLysThrSerValIleLysAsp... 1583
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3660 ACACCTACCGTTCGCAAGATTTCGCGCTACCGCAACAAACGACCTGC 3709
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|

```

```

1584 ....HisAsnGlyLysLeuPheThrGly...LysGluLeuLysAspAlaA 1598
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3710 GCCAAATCGGTATGCAG.....AAAACCTCGCGCAGC 3741
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1598 snAsnAsnGlyValThrValThrGluThrAspGlyLysAspGluLysAsn 1614
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3742 GGGCGCGTC.....GGCATCTGTTTTCGCACACCGG..... 3774
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1615 GlyLeuValThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTr 1631
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3774 ..... 3774
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1631 parValLysThrThrGlyAlaAsnGlyLysAsnAspAspPheAlaThrV 1648
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3775 .....ACGAAACACCTTCGACGCGCATCGCACTCGCA 3813
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1648 alaLaserGlyThrAsnValThrPheAlaAspGlyAsnGlyThrAla 1664
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3814 CGGCTTGGC.....CAGCGCGCGTTTCGGGCAATACGCATCGA 3854
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1665 GluValThrLysAlaAsnAspGlySerIleThrValLysTyraVal... 1680
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3855 CAGTTCCTACATCGCATCAGCGCGCGCGGCTTTACACGCGCAGCC 3904
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1681 .....LysV 1682
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
3905 TTTCAGACGCGATC.....GGAGCAAAATC 3930
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
1682 alaLAspGlyLeuLysLeuAspGlyAspLysIle 1693
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-47
seq_documentation_block:
; Sequence 47, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 2354
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-47
alignment_scores:
      Quality: 229.50      Length: 1495
      Ratio: 0.339      Gaps: 72
      Percent Similarity: 45.284      Percent Identity: 19.130
alignment_block:
US-09-303-518d-649 x US-09-268-347-47 ..
Align seg 1/1 to: US-09-268-347-47 from: 1 to: 2354
172 GCGGAATAAAGGCAAGTTTCAGTGGCGGGAAGAAAGATATGACGTTTA 221
      |||::: |||::|::|::|::|::|::|::|::|::|::|
436 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGluIvaL... 451
      |||::: |||::|::|::|::|::|::|::|::|::|::|
222 CAACAAGAAAGGAGGAGTTGGTC.....GGCAATCAATG...A 256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
452 ...LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 467
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
257 CAAGACCCCGATGATGATTTTCTGTGCTGCTGCGGTAACGCGCTGGCG 306
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
467 allYsGlnAspGlyAlaAsnPheThrItyrSerLeuGlnAspAlaLeuThr 483
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|
307 GCATTGCTGGC..... 318
      ::: |||::: |||::|::|::|::|::|::|::|::|::|::|

```

```
484 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 500
      ::|||::
319 .....GATCAATATATTGACGGTGGCACAATACG 349
      |||
500 alysthrValIleasnlyAspGlyLeuThrIleThrProIleGlyAsn 517
      |||
350 GCGGCTTAACAACGTTGATTGGCGGAGAGAAATCCCGATCA 399
      |||
517 Ly.....GlyThrThrGlyThrAsn..... 523
400 CATCGTTTACTTAAATTTGAAACGGGAATATTAAACAGGAC 449
      |||
524 .....ThrIleSerValThrLys..AspGlyIleLysAlaGlyAs 536
450 TAAAGCCATCTTATGCGCGCATATATCCCGCTTGCAATAA 499
      |||
536 nlySAlA..... 538
500 TTGTACACAGATGCAAACTGTTGAATGACAGTTATATGATGGCG 549
539 ..... 546
550 AAATATATCGATCAAAATAATTAACCTGACCGTTCGATTTGGGACG 599
      |||
547 ArgAlaThrAspAspAlaAsnPhe..... 554
600 CAGGCAATATTGGCGATGTGATGAAGTGAAGCCCAATAC..... 643
555 .....AspValLeuAsnAsnSerAlaThr 563
644 AAATTCATATCATATTGCAAGTGGCTATCTGGCTGTTGGTGCAAT 693
      |||
563 spleuAsnAlaGlnHisValGluAspAlaThrLys..... 573
694 ACCTTTGCAAAATGGATCAGGTGGTGACAGCTCAACTAGTAGTGA 743
574 .....GlyLeuLeuAsnLeuAsnGly 581
744 AAAAATTAAACATAGCCCATATGTTTTCACACAGAGGCTCATTTG 793
      |||
581 sAsnAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrVal 598
794 GCGACAGTGGCTCACCAATGTTATGATGTCACCAAAAGCAAAAGTG 843
      |||
598 Ly.....AspLeuArgLysLeuGlyTyr 605
844 TTAATT.....AATGGGATATTGCAAAAGGCAACCCCTATATAG 884
      |||
606 ValValSerThrLysAsnGlyThrLysGluGluSerAsn..... 618
885 AAAAAGCAATGGCTTCACAGTGGTGTGTAAGATTGGTTCTATGATGAA 934
619 .....GlnValLysGlnAlaAsp.....GluVal 627
935 TCCTTGGAGATACCATTCAGTATTCACAGACGTCACAAATGG 984
      |||
627 euphethrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 643
985 AAATACTCTTTTAAAGCAGATATATATGCAAGAAAAATCAATGCCAA 1034
      |||
644 LysHisThr..... 646
1035 ACATGAACACAAATTCTGCTAATAGATTAAAAACAGAACGGTTCAAT 1084
646 ..... 646
1085 TGTTAATGTTTCTTATCGAGACAGACAGAACCTGTTATCATGCT 1134
      |||
647 ..IleThrValSerValAlaGluThrLysAlaAspGlyLeuGlyLys 662
1135 GCAGGTGCTCAACAGCTTATCGACCAAGACTGAAATATGAGAAAAATAT 1184
      |||
```

1851 AGACGCGACCCGTCGCTTCGCGGGGAACA.....AATT 1885
971 pThrGlyThrGlyLeuValThrIalaIysThrValIleAspIalaValAsnL 988
1886 TAAACGCGCAACATCAAGCAAAACGCGCAACTGTTTTCAGCGGAGA 1935
988 ySerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyAla 1004
1936 CCAACACCGCAGCGCTACATCATTTAAACGCACTTGTCGGCAAAAAGA 1985
1005 ThrAlaValAsnAlaGlyAsn.....AlaG 1013
1986 GGGCATTCCTCGCGGGGAATCGTGTGGACACAGCATGATCAACGCCA 2035
1013 uThrValThrSerGlyThrSerVal..... 1021
2036 CATTTAAGCGGAAAACCTTCGAAATTTAAAGCGGACAGCGCGTGCTTCC 2085
1022AsnPe.....LysAsnGlyAsnAlaThrThrAla 1031
2086 CGCAATGTTGCCAAAGTGAAGAGCGCATTTGGCATTTGACCAATCAAGCCCA 2135
1032 ...ThrValSerLysAspAsnGlyAsnIleAsnValLysTrpAspValAs 1047
2136 AGCAAGTTTGGTGTGCGACCGCATCAAAAGCCACAACTGTACAGCTT 2185
1047 nValGlyAspGlyLeuLysIleGlyAspAspLysIleVal.....A 1062
2186 CGGACGTGAGCGGCTGACAAATTTGTGCGAAAAACCATTAACCGAGCAT 2235
1062 IaAspThrThrThrLeu.....ThrValThrGlyGly 1072
2236 AAAGTATTGCTTCATTGACTAGACGCGACATCAAGCGGCAATGTGATCT 2285
1073 LysValSerValProIaGlyAlaIAsnSerValAsnAsnAsnLysLysLe 1089
2286 TGGCGATCAACGCTCATTTAAATCTCAACAGCGCTTGCCACACTCAAGGCA 2335
1089 u.....ValAsnIaGluGlyLeuAlaThrAlaLeuAsn 1101
2336 ATCTTAGT.....GCNAATGGCGATACAGT 2361
1101 snLeuSerTrpThrAlaLysAlaAspLysTrpAlaAspGlyLysSerGlu 1117
2362 TAATACGTAGCCACACAGCCACCCAAAAGCGCAACTTAGCTTC..... 2406
1118 GlyLThrAspGlnGluValLysAlaGlyLysValThrPheLysAl 1134
2407GTGGCAATGCCCAAGCAACATTTAAT..... 2433
1134 agLysLysAsnLeuLysValLysGlnSerGlnLysAspPheThrLysSerL 1151
2434 ..CAAGCCACATTTAAAGCGGCAACACATCGGCTCG.....GGCAATGCT 2475
1151 euLAsnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyLThrAla 1167
2476 TCAATTAACTAAGCAGACCGCCTTACAAAAGCG.....AGTCT 2516
1168 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1184
2517 GAGCGCTTTCGGCAACGCTAAGGCAAGCAAGCATTCGGACATCAAGC 2566
1184 eThrLeuAlaAsnGlyAlaIalaIaGlyLThr.....AspAlaSerAsnG 1199
2567 GTAAT.....GTCTCCCTAGCGCGATTAAGCA 2592
1199 LysAsnThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1215
2593 GTATTCATTTTGAAGACAGCGGCTTACCGGACAAATCAACGGCGGCCAA 2642
1216 lIeThrAsnValLysSerAlaLeuLysThr.....TyrLysAspThrG 1230

2643 GGATAGGGCATTAACCTTAAGACAGCGGAATG.....ACGTCGGCT 2686
1230 masThrIalaAspLutThrGlnAspLysGluPheHsalaIaValLysA 1247
2687 CAGGACAGGAATTAAGCAATTTAAACCTTGACAGACCGCACTTACA... 2733
1247 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1263
2734 ..CTCAATTCCGCTATCGCCAC.....GATCGCGACAGGGC 2768
1264 LysThrAspAsnAsnGlyLysHsIsthValThrIleAspValAlaGluAl 1280
2769 GCAACCGCGCAGTGC.....ACAGATGCG.....CCGCGCC 2800
1280 aLysValGlyAspGlyLeuGlnLysAspThrAspLysIleLysLeuL 1297
2801 GCCGTTGGCCCGCTGGCCGCGCTTCCTTATTCGCTT.....ACACCG 2844
1297 ysValaAspAsnThrAspLysAlaAsnLeuLeuThrValaAspAlaThrLys 1313
2845 CCAACTTCGCTA...GAATCCGTTTCAACACGCTGACG..... 2880
1314 GlyLalaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaTh 1330
2881GTAAACGGCAATTTGACGCTCAGGGAACATTCGCT 2917
1330 rThrAlaGlnGlyThrAsnAlaAsnGluArgLysValaValaValLysG 1347
2918 TTATGTCGGAACCTTCGCTACCGCGACGACAAATGACATGCGCGGA 2967
1347 LysSerAsnGlyAlaThrAlaThrGluThrAspLysLysVala... 1362
2968 AGTTCGGAAGCACCTTACACCTTGCGGCTCAACAATACCGCAAGAAC 3017
1363 ...ThrValGlyAspAlaAlaLysAlaIleAsnAspAlaAlaThr.... 1376
3018 TGCAGCGCTGAGCAATTTGACGCTAGTGGACAGAAACAAACAAACCGC 3067
1377PheValLysValGluLysAsnAspSerAlaThrI 1388
3068 TGTCGGAACCTTAATTTACACCTTGCAAAAGCAACAGATGATCGCGC 3117
1388 lAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAlaGly 1404
3118GGTGGCGCTTACCAACTCATCCGCAAGACGC 3149
1405 AspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspG 1421
3150 CGAG.....TTCGGCTCGCATAT.....CCGGTCAAGAGCAAG 3184
1421 LysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSerAla 1438
3185 AGCTTTCGCAAAATC.....GGCAAGCGCAAAACC 3216
1438 hrValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValaSnIle 1454
3217 AAAAAACGGCGGAA.....AAGCAACGGCGCAAAAGCT 3251
1455 ThrSerAspThrLysGlyLeuLysPheAlaLysAspSerLysTrpGlyAs 1471
3252 TGACGGCGCTGATTTGCGGCGCGCGCGGATGCGCTCGAAAGAAGCAAGCG 3301
1471 pAspAlaAsnIleHsIleAsnGlyIleAlaSerThrLeuThrAspThrL 1488
3302 TT.....GCCAACCAGCGCCGCGCAGGACAGCGGCGGAAATATGCGCAT 3345
1488 euLeuAsnSerGlyAlaThrThrAsnLeuGlyLysAsnGlyLe..... 1502
3346 ATGACGGCGGAGGAGGAGAAAGAGGGGTGCGAGCGGATTAAGACACCGC 3395
1503ThraspAsnGluLysLysArgAlaAlaSerValLysAspValLe 1517
3396 CTTGGCGAAGAACCGCGGAACCCGCGCGCTACACCGCCCTTCC 3445

```

1517 uasnlaaglytrpasmvalargylvalylsprolaser..... 1530
1530 .....
3446 CCCGGCCCGCGCGCCGGGATTTGCCGCAACTGCACCCCAACCG 3495
1530 .....
3496 CAGCCCAACCGCAGCGGCGGCTGATCAGCGGTTATGCCAATAGCGGTTT 3545
1531 .....
3546 GAGT.....GATTTTCCGCGCAGCTCAACAGGTTTCCCGCTACAGG 3589
1535 IGLuasnleasphnevalaiathrtyraspthrvalasphnevalserg 1552
3590 ACGAATTAGAC.....CGCGTATTGGCGAAGACCGCGGCAAC..... 3627
1552 LysAspLysAspThrThrSerValThrValGluSerLysAspAsnGlyLys 1568
3628 .....GCCGTTTGACAAGCGGCGATCCGGACACCA 3659
1569 ArgThrGluValLysIleGlyAlaLysThrSerValIleLysAsp..... 1583
3660 ACACACGCTTGGCAAGATTCCCGGCTACCGGCAACCAACGACCTGC 3709
1584 .....HisasnGlyLysLeuPheThrGly...LysGluLeuLysAspAla 1598
3710 GCCAAATCGGTATGCG.....AAAACCTCGGCGAGC 3741
1598 snAsnAsnGlyValThrValThrGluThrAspGlyLysAspGluGlyAsn 1614
3742 GCGCGCGCTC.....GGCATCCTGTTTGCACACACCGG..... 3774
1615 GlyLeuValThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTr 1631
3774 ..... 3774
1631 PArgValLysThrThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrV 1648
3775 .....ACGAAACACCTTCGACGACGCGATCGGCACTCGCA 3813
1648 alaIaSerGlyThrAsnValThrPheAlaAspGlyAsnGlyThrThAla 1664
3814 CGGCTGCG.....CACGGCGCGTTTTCGGAATACGCGCATCGA 3854
1665 GluValThrLysAlaAsnAspGlySerIleThrValLysTyraVal.. 1680
3855 CAGGTTCTACATCGCATCAGCGGCGGCGGTTTACGACGCGGACGC 3904
1681 .....LysV 1682
3905 TTTGACAGCGGCATC.....GGAGCAAAATC 3930
1682 alaIaAspGlyLeuLysLeuAspGlyLysPylsile 1693
seq_name: /cgn2_6/ptodata/1/laa/6B.COMB.pep:US-08-968-685A-10
seq_documentation_block:
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSTIA, LAURA
; TITLE OF INVENTION: MORAXIELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USPS THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968, 685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-968-685A-10

```

```

alignment_scores:
  Quality: 223.00      Length: 1548
  Ratio: 0.318        Gaps: 71
  Percent Similarity: 45.349   Percent Identity: 18.992

```

```

alignment_block:
US-09-303-518D-649 x US-08-968-685A-10 ..

```

```

Align seg 1/1 to: US-08-968-685A-10 from: 1 to: 2123
469 GCGGATATCATATATGCGCGGTTTGCATTAATTGTCCACAGATGCAGAAC 518
779 GlyAspLeuThrPheSerLysGlnAsnThrGlyAlaThrProAlaThrAs 795
519 TGTGTAATGACCGATTATATGATGCGGCAATATATATCATCAAAAT. 567
795 nserLysThrIleTyrlGlyValAspGlyLeuLysPheThrAspAsnG 812
568 .....AATTACCTGACCGCTTGTGATTGGGGA 597
812 LylleAlaLeuAspGlyThrThrTyrlleThrLysAspLysValGlyPhe 828
598 GCGAGCATATATGCGCATCTGATGAGATGAGACCAATACCGGGAAG 647
829 AlaLysGlnAspGlySerLeuAspLysSerLysProTyrlLeuAspLysAs 845
648 TTCATATCATATAT..... 660
845 PylsLeuLysValGlyGluValGluIleThrThrAsnGlyIleAsnAlag 862
661 ..GCAAGTCGCTATCTTGTGCTCGTTGCGCAATACCTTGACAAAT 708
862 LylLysAlaIleThrGlyLeuSerAsnThrLeuThrAspAlaThrAsn 878
709 GCATGAGTGTGCGACATCACTTAGT.....AGTGAANA 746
879 AlaThrThrGlyHisValThrGlnLeuGlyIleValAspSerThrAspLys 895
747 AATTAAACATACCCATATATGTTTTCACCAACAGAGGC..... 786
895 sthrArgAlaIaSerIleGlyAspValLeuAsnIaGlyPheAsnLeul 912
787 ..TCATTGCGCAGAGTGTGCTACCAATCTTATATATAGATGCCAAAG 834
912 yasnAsnGlyAspAlaLysAspPheValSerThrTyrlAspThrValasp 928

```

835 CAAAAGCTTAATTAATGGGCTATTGCAACGGCGAAC...CCCTATAT 881
:::|||||
929PheileasnIalathrAlalalysValthrTyrAs 942
882 AGGAAAGCAATGGCTCCAGCTGGTGTAAAGATGGTGTATGATGAG 931
|||||:::
942 pGlyLysAlaSerLysValAla.....TyrAspY 952
932 AAATCTTGGTGGAGATACC...CATTCAGTATTCAGAACCCAGCTCA 978
||| |||
952 alasnValaspGlyThrThrLleHisLeuthrGlValalaspGlyasnLys 968
979 AAT.....CG 983
|||
969 AsnGlnIleGlyValLysThrThrLleuthrLysThrAspAlalaspG 985
984 GAATATCTCTTTAAC...GACGATATATATGGCACAGCAAA.....A 1024
| : : : : :
985 yaSPlySAlaIleasnPheserValasnSerGlyaspSPlySAlaLeuI 1002
1025 TCATGCCAAACATGAAACACATTCCTGCTATAGATTAAACAGCA 1074
|||||
1002 leasnAlalys.....AspIlealaspasnLeuasnThrLeu 1014
1075 ACCGTTCAATGTTTAACTTT...TCTTTATCCGAGACAGCAAGAGAAC 1121
::: : : : : :
1015 AlagIyGlnIlethrValasnThrLysGlyThrAlaspThrAlaleuIthr 1031
1122 TGTTATCATGCTGCAGCTGTGTCAACAGTATTCGACCACAGCTGAATA 1171
1031 rPheGln.....ValLysLysValLysGlnAsnGlyAspA 1043
1172 ATGAGAAATATTTCTTATTGACGAGAAAGGCGAATTGATCTT 1221
: : : : :
1043 spaSPasnaspAlaspThrLlethrValGlyLysasp.....Alalys 1057
1222 ACCGACACATCAATCAAGGTGCTGAGCATTAATTTCCAGAGATTT 1271
||| : : : : :
1058 ThrasnGlnValasnThr.....LeuLysLeuLysGlyLys.. 1069
1272 TACGTCCTCCCTGAAATAACGAACTGGCAAGCGCGGCTTCATA 1321
1070AsnGlyLeuaspIleGln 1076
1322 TCAGTGAAGACAGTACCGTTACTTGAAGATAACGGCGTGCAGAACGAC 1371
: : : : :
1076 hrAsnLysaspGlyThrValThrPheGlyLleasn..... 1087
1372 CGCGTCGCAAAATCGGCAAAAGCAGCTGCACGTTCAACGCAAAAGCGA 1421
1087 1087
1422 AAACCAAGGCTCATCAGCGTGGCGAGGTACAGTATTTGGATCAGC 1471
: : : : :
1088 .ThrGlnSerGlyLeuLysAlaGlyAsnasnThrThrLeuAsnAsnAsn 1104
1472 AGGCAAGACGATAAGCGCAAAACAGCCCTTACTGAATCGGCTTGCTC 1521
: : : : :
1104 LysLeuSerIleLysasn.....Thr 1110
1522 AGGGCGAGGGGTACGGTCAACGATGCGGATATCAGTTCAACCCCGA 1571
: : : : :
1111 AlagLysasnGlnGlnIleGlnValGlyAlasp..... 1121
1572 CAAACTCATTTCCGCTTCCGCGGAGCGATTGATTTAAACGGGCAAT 1621
1122G 1122
1622 CGCTTGGTTCACCGTATTTCAAAATACGATGAGGGCGGATGATGTC 1671
: : : : :
1122 LysValLysPheAlalalysValasnGlyValValGlyAlaGlyLle... 1137
1672 AACCAATCAAGACAAAGATCCACCGTTACATTACAGGCATATAAGA 1721

1138AspGlyThrThrArgIleThr...ArgaspG 1147
1722 TATTCCTAACCGCGCAATTAACACAGCTTGAT...AGCAAAAAAGAA 1768
: : : : :
1147 uIleGlyPheAlaGlyThrAsnGlySerLeuAspLysSerLysProIle 1164
1769 TTGCTTCAACGGT.....TGCTTTGGCGAAAGATAGCACCMAAGC 1812
1164 euSerLysaspGlyLleasnAlaGlyGlyLysLysIleThrAsnIleGln 1180
1813 AACGGCGGCTCAACCTGTTTACCAGCCCGCCAGAAAGACCCAGCT 1862
: : : : :
1181 SerGlyGlnIle.....AlaGlnAsnSerAsnAs 1190
1863 GCTGCTTCCGCGCA.....ACAAATTTAAACGGCA 1894
: : : : :
1190 palValalThrGlyGlyLysIleTyrAspLeuLysThrGlnleuGlnsl 1207
1895 ACATCAGCGCAACA.....AACGCAAACTGTTTTCAGC 1929
1207 yaIleSerSerThrAlalysThrAlaGlnasnSerLeuHisGlnPheSer 1223
1930 GGCAGACCAACACCGCACGCTACATCATTTAAACGACATTCGTCGA 1979
1224 ValAlaaspGlnGlnGlyAsnAsnPheThrValSerAsnProLysSer 1240
1980 AAAAGAGGCGATTCCTCGCGGGAATCGTGTGGACACAGCTGATCA 2029
: : : : :
1240 rTyrAspThrSerLysThrSerAspValIle..... 1250
2030 ACCGCACTTTAAACGGCAAAAC...TTCCAAATTTAAAGCGGACAGCG 2076
1251ThrPheAlaGlyLysasnGlyIleThrThrLysValasnLysGly 1265
2077 GTGTTTCCCGCAATGTGCCAAAGTGAAGCGATTTGGCATTTG.... 2121
1266 ValValalargValGlyLleaspGlnThrLysGly.....LeuThrTh 1279
2122ACCAATCAGCGCCAGCAGCTTTTGTGTGCG 2152
1279 rProLysLeuThrValGlyAsnAsnAsnGlyLysGlyLleValIleAsps 1296
2153 CACCGCATCAAGCCACACCAATCTGTACAGCTGCAGCTGCGAGGCTG 2202
1296 ergLAsnGlyGlnAsnThrIle.....ThrGlyLeu 1306
2203 ACAAAATGTGTGAA.....AAACCAT 2225
1307 SerAsnThrLeuAlaAsnValThrAsnAspLysGlySerValArgThrTh 1323
2226 TACCGACGATAAGATGCTTCTCATTAAGCCGACATCAGCGCA 2275
: : : : :
1323 rGlnGlnGlyLysIleLleLysaspGlyAspLysThrArgAlaAlaSerI 1340
2276 ATGTGATCTTCCGATACAGCTCATTTAAATCTCAGAGGCTGCCCA 2325
1340 leValasp.....ValLeuSer 1345
2326 CTCACGCGCAATCTTAGTCAAAATGCGATACACGTTATAGTCAAGCA 2375
: : : : :
1346 AlagLysPheAsnLeuGlnGlyAsnGlyLysAlaValaAspPheValSerTh 1362
2376 CAACGCCACCCCAAAAC.....GCAACCTTAGCTCGGGGCAATG 2416
1362 rTyrAspThrValasnPheAlaaspGlyAsnAlaThr.....Thra 1376
2417 CCCAAGCAATTTAATCAAGCC..... 2439
1376 lalysValThrThrAspAspThrSerLysThrSerLysValValLysrAsp 1392
2439 2439


```

1393 ValAsnValAspAspThrThrIleGluValLysAspLysLysLeuGlyVal 1409
2440 ...ACATTAAACGGCACATCGCTTCGGCAATGCTTCAATTATTC 2485
1409 LysThrThrThrThrLeuThrSerThrGlyThrGlyAlaAsnLysPheAla 1426
2486 TAGAGGACACCGCGGTACAAAACGGCAGTCTG 2517
1426 euserAsnGlnAlaThrGlyAspAlaLeuValLysAlaSerAspIleVal 1442
2518 ...ACGCTTTCGGGAC 2532
1443 AlaHisLeuAsnThrLeuSerGlyAspIleGlnThrAlaLysGlyAlaLeu 1459
2533 ...GCTAAGCAACAGTAAGCCATTCCGCACTCAAGCGTAATGCTCC 2578
1459 rGlnAlaAsnSerSerAlaGlyTyrValAspAlaAspLysAsnLysValI 1476
2579 TAGCCGATTAAGGACGATTAATTCATTGTAAGCGCGCTTACCGGACAA 2628
1476 LetyrAspSerThrAspAsnLysTyrTyrGlnAlaLysAsnAspLysThr 1492
2629 ATCAGCGCGCGCAAGGATACGCAATTACATTAAAGACAGCAATGGAC 2678
1493 ValAspLysThrLysGluValAlaLysAspLysLeuValAlaGlnAlaG 1509
2679 GCTGCCGCGACGACGAAATTAGCAATTAACTTGACACGCGCACCA 2728
1509 nThrProAspGlyThr...LeuAlaGlnMetAlaAsnValLysSer 1522
2729 TTACACTCAATTCGCGCTATCGCCACGATGGCGGCGGCA 2772
1523 ...ValIleAsnLysGlnGlnValAlaAsnAspAlaAsnLysLysGlnGlyLe 1538
2773 ...ACGGCAGTGGCAGCAGATGC 2792
1539 AsnGluAspAsnAlaPheValLysGlyLeuGlnLysAlaAlaSerAspAs 1555
2793 GCCCGCGCGCGCTTCGCGCGCGCTTCCTATTATTCGCT 2838
1555 nLysThrLysAsnAlaValAlaValThrValGlyAspLeuAsnAlaValAlaG 1572
2839 ...ACACCG 2865
1572 InThrProLeuThrPheAlaGlyAspThrGlyThrThrAlaLysLeu 1588
2866 TTCACACGCGTACGCGTAACGCAATTGAACGCTCAGCAACATTCG 2915
1589 GlyLeuThrLeuThrIleLysGly...GlyGln 1598
2916 CTTATATGTCGGAAGCTTCGCGTACCGAGCGACAAATTGAAGCTGGCG 2965
1599 ...ThrAspThrAsnLysLeuThra 1606
2966 AAAGTTCGAGGCACTTACACCTGGCGGTCAACAATTACCGGACGAA 3015
1606 sPAsnAsnIleGlyValVal 1612
3016 CCTCAAGCGCTCGAACAATTGACGTAGTGAGAAAAGCAACAAC 3065
1613 ...AlaGlyThrAspGlyPheThrValLysLeuAlaLysAsp 1625
3066 GCTCTCCGAAAACCTTAATTACCGTCAAAAACGACAGCTGATGCG 3115
1626 ...LeuThrAsnLeuAsn...SerValAsnAlaG 1635
3116 GCGGCTGGCGT 3150
1635 LysGlyThrLysIleAspAspLysGlyValSerPheValAspSerSergly 1651
3151 GAGTTCGCGCTCATTAATCCGGTCAAGAACAGCTTCCGCAACACT 3200
1652 GlnAlaLysAlaAsnThrProValLeuSerAlaAsnLysLeuAspLeuG 1668
3201 CGGCAAG 3235
1668 yGlyLysValIleSerAsnValGlyLysGlyThrLysAspThrAspAla 1685
3236 ACAACGGCAAGCGCTTGAC 3267
1685 lAsnValGlnGlnLeuAsnGluValAlaArgAsnLeuLeuGlyLeuGlyAsn 1701
3268 GCCGGCGCGATGCGCTGAAAGACAGAAAGCGTTCCGCAACGGCGCG 3317
1702 AlaGlyAsnAsp 1705
3318 GCAGCGAGCGGGGAAATATCGGCATTATGACGCGGAGAAAG 3363
1706 AsnAlaAspLysAsnGlnValAsnIleAlaAspIleLysAspPro 1722
3364 ...AAAAACGGGTGCGAGCGGATTAAGACACCGCGCTTG 3399
1722 snSerGlySerSerSerAsnArgThrValIleLysAlaGlyThrValLeu 1738
3400 GCGAACAAGCGCGCAAGCGCAACCGCGCGCTACGAC 3438
1739 GlyGlyLysGlyAsnAsnAspThrGlnLysLeuAlaThrGlyGlyValG 1755
3439 ...GCCCTTCCCGCGCGCGCGCGCGCGCGGATTTGCCCAACTGC 3484
1755 nValGlyValAspLysAspLysAlaAsnGlnLysAspLeuSerAsnVal 1772
3485 AACCCCAACCGCACGCCCA 3531
1772 rPalLysThrGlnLysAspGlySerLysLysAlaLeuLeuAlaThrTyr 1788
3532 GCCAATAGCGGTTGAATGTAATTTCCGCGCACGCTCAACGCGTTTCC 3581
1789 AsnAlaAlaGlyGlnThrAsnTyr...LeuThrAsnAspProAlaGln 1804
3582 CGTACAGAGAGAAATTAGACCGCTATTGCGCA 3615
1804 alle 3615
3616 ...GACCGCGCAACGCGCT 3633
1817 lAsValAsnAspGlyAsnGlnGlnProValValGlnLysArgAsnGlyLe 1833
3634 TGGACAAAGCGCGATCCGGACACCAACACTACCGTTCCCAAGATTTCCG 3683
1834 AspSerSerAla...SergLysHisSerValAlaIleGlyPheG 1848
3684 CGCCTACCGCGCAACAACGACCTGCGCAAAATCGGTATGCAAGAAAAC 3733
1848 nAlaLysAlaAspGlyGlnAlaAlaValAlaIleGlyArgGlnThrGln 1865
3734 TCGCAGCGCGCGCTCGGCAATCTGTTCCACAAACCGGAC...GAA 3780
1865 lAsGlnSerIleAlaIleGlyAspAlaGlnAlaThrGlyAsp 1881
3781 AAACCTTGCAGACGCGCATCGCAACTGCGCAAGGCTTCCGACGGCG 3830
1882 GlnSerIleAlaIleGlyThrGlyAsnValAlaIleGlyLysHisSerG 1898
3831 CGTTTCGGGCAATACGCAATGACAGCTTCAATCGGCATCAGCGCG 3880
1898 yAlaIleGlyAspProSerThrValLysAlaAspAsnSerTyrSeraI 1915
3881 GCGCGGT 3918
1915 LysAsnAsnAsnGlnPheThrAspAlaThrGlnThrAspValPheGlyVal 1931
3919 GGAGGCAAAATCCG 3933
1932 GlyAsnAsnIleThrValThrGlnSerAsnSerValAlaLeuGlySera 1948

```

```

3934 .....CGCCGGTGTGCTGATTCAGCGCATTCAGCGCATTCAGCGG 3973
1948 nserAlleSerValaglyThrHisalaglyThrHisalaglyLysSerA 1965
3974 CCGGTTTCGGC..... 3984
1965 spgLyThrAlaglyThrThrThrAlaglyAlaThrGlyThrValys 1981
3985 GGATTCGGCATTCAGCGCATTCAGCGCGCGCATTCGTCACAAA 4034
1982 GlyPheAlaglyThrAlaValGlyAlaValSerValGlyAlaSerG 1998
4035 AGCGATTCAGCGCATTCAGCGCATTCAGCGCATTCAGCGCATTC 4065
1998 yAlaGlyArgArgIleGlnHisnValAlaAlaGlyGluValSerAlaThrs 2015
4066 .....GCCACCCCGCGCTT 4080
2015 erThrAspAlaValasnglySerGlnLeuTyrlsAlaThrGlnGlyIle 2031
4081 GCA.....TTCAACG 4091
2032 AlaAsnAlaThrAsnglyLeuAspHisArgIleHisGlnAsnglyLys 2048
4092 CTACCGCGCGCGCATTCAGCGCATTCAGCGCATTCAGCGCGCAAC 4141
2048 sAlaAsnAlaGlyIleSerSer.....AlaMetAla 2059
4142 TTTCATCAGCGCTTATTCAGCGCTTATTCAGCGCGCTTATTCAG 4191
2059 etAlaSerMetProGlnAlaTyrlleProGlyArgSerMetValThrGly 2075
4192 AAATGCGCAACGCGCATTCAGCGCGCATTCAGCGCGCATTCAGCG 4241
2076 GlyIleAlaThrHisAsnglyGlnGlyAlaValAlaValGlyLeuSerly 2092
4242 A..ACCGCGATGCGGAAATG.....GCCGTAAACCGCGGAAA 4276
2092 sLeuSerAspAsnglyGlnThrValPheLysIleAsnglySerAlaSpr 2109
4277 TCAAGGTTTCAGCGCTGTCCTCCAGCGCGCGCGCGCGCGCGCG 4320
2109 hArgGly.....HisValGlyAlaAlaValGly 2118

seq_name: /cgn2_6/plodata/1/1aa/5A_COMB.pep:US-08-409-995-4
seq_documentation block:
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Honbach, Test, Albrighton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409/995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304

```

```

; REFERENCE/DOCKET NUMBER: A-61053/RPT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-4

alignment_scores:
  Quality: 218.50      Length: 1492
  Ratio: 0.322         Gaps: 73
  Percent Similarity: 45.442   Percent Identity: 19.035

alignment_block:
US-09-303-518d-649 x US-08-409-995-4 ..

Align seg 1/1 to: US-08-409-995-4 from: 1 to: 1912

172 GCGCAAAATTAAGCGCACTTTCAGTCGCGCGCAAAAGATATTCAGCTTA 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 AlaAspThrAspGlyAlaLeuGlnGlyIleSerLysAspGlnGluVal.. 450
222 CAACAAAAAGGAGGAGTTGTC.....GCCAATCAATG...A 256
||| ||||| ||| |||||:|||||:|||||:|||||:
451 ...LysAlaGlyGlnThrValThrPheLysAlaGlyLysAsnLeuLysV 466
257 CAAAACCCCGATGATTTTCGTGCTGCGGTACGCGCGCGCGCGG 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 allysGlnAspGlyAlaAsnPhetThrTyrlSerLeuGlnAspAlaLeuThr 482
307 GCATGCTGCGC..... 318
483 GlyLeuThrSerIleThrLeuGlyGlyThrHisAsnglyLysAsnAspAl 499
319 .....GATCAATATATTTGAGCGTGCACATTAACG 349
499 alysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 516
350 GCGGCTATACACGTTGATTTGGTGGGAGAGAAATTCACATGAA 399
||| ||||| ||| |||||:|||||:|||||:|||||:
516 ly.....GlyThrThrGlyThrAsn..... 522
400 CATCGTTTACTTATTAATTTGGAACGGAATATTTAAAGCAGGAC 449
|||:|||||:|||||:|||||:|||||:|||||:|||||:
523 .....ThrIleSerValThrLys..AspGlyIleLysAlaGlyLys 535
450 TAAAGCCATCTTATGCGCGGATTAATGCGCGGTTTCATTAAT 499
|||||:|||||:|||||:|||||:|||||:|||||:
535 nLysAla..... 537
538 .....IleThrAsnValAlaSerGlyLeu 545
500 TTGTCAACATGCAGACCTGTTGAATGACCATTTATATGATGGGCGG 549
538 ..... 545
550 AAATATATGATCAAAATATTAACCTGACCGTTCGATTTGGGCGAG 599
|||||:|||||:|||||:|||||:|||||:|||||:
546 ArgAlaTyrlAspAspAlaAsnPhe..... 553
600 CAGGCAATATGCGCATTCGATGAGATGAGCCCAATAC.....CGCG 643
554 .....AspValLeuAsnAsnSerAlaThrA 562
644 AAAGTCATATCATATTCAGATGCGTATTCGTCGTCGTCGTCGAT 693
|||:|||||:|||||:|||||:|||||:|||||:|||||:
562 spleuAsnArgHisValGluAspAlaTyrlLys..... 572
694 ACCTTGCAACAAATGATCAGTGTGTGCGACAGTCACATTAAGTAGTAGA 743

```


2286 TGGCGATCAGCCTGATTTAAATCTCAGGGCTTGGCAACCTGACGGCA 2335
 1088 u.....ValasnaIaIuGIuLeuAlaThrAlaLeuAsna 1100
 2336 ATCTAGT.....GCAANTGGGATACACGT 2361
 1100 snLeuSerThrAlaLysAlaAspLysTyrAlaAspGlyLysSerGlu 1116
 2362 TATACAGTCAGCCACACGCCCAACCAACCACTTACGCTC..... 2406
 1117 GtGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAl 1133
 2407GTGGCAATGCCCAAGCAACATTTAAT..... 2433
 1133 aGlyAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1150
 2434 ..CAAGCCACATTTAAACGGCAACACATCGGCTCG.....GGCAATGCT 2475
 1150 euGlnAspThrLeuThrGlyLeuThrSerLLeuThrLeuGlyLysAla 1166
 2476 TCATTTAATCTAAGGACACGCCGCTACAAACAGGC.....AGTCT 2516
 1167 AsnGlyAlaGlnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1183
 2517 GACGCTTTCGGGCAACGCTAAGCAACAGTAAAGCATTCCGACCTCAACG 2566
 1183 eThrLeuAlaAsnGlyAlaAlaIaIaGlyThr.....AspLaserAsnG 1198
 2567 GTAAT.....GTCTCCCTAAGCCGATTAAGCA 2592
 1198 LysSerThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1214
 2593 GTATTCATTTTGAAGCAGCCGCTTACCGCAATCAACGCGCGCA 2642
 1215 IleThrAsnValLysSerAlaLeuLysThr.....TyrLysAspThrGlu 1229
 2643 GGATTCGGCATTTACACTTAAAGACAGCAATGG.....ACGCTGCGCT 2686
 1229 nasThrLaserGluThrGlnAspLysGluPheIleAlaValLysAs 1246
 2687 CAGGACGGATTTAGGCAATTTAACTTGACAGCCCACTACA... 2733
 1246 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1262
 2734 ..CTCAATTCGGCTATCGCCAC.....GATGCGCAGGGCG 2768
 1263 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1279
 2769 CCAACCGCGCAGTGGC.....ACAGATGG.....CCGCGCC 2800
 1279 alyValGlyAspGlyLeuGluLysAspThrAspGlyLysIleLysLeu 1296
 2801 GCCGTCGCGCGCTTCCGCGCCCTCATATATCGCTT.....ACACGG 2844
 1296 yValAspAsnThrAspGlyAsnAsnLeuThrValAspAlaThrLys 1312
 2845 CCAACTTCGGTA...GAATCCGTTTCAACAGCCTGACG..... 2880
 1313 GlyLaserValAlaLysGlyLysPheAsnAlaValThrThrAspAlaThr 1329
 2881GTAACGGCAATTTGAACGGTCAGGGAACATTCGGCT 2917
 1329 rThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValLysG 1346
 2918 TTATTCGGAACCTCTCGGTCAGCGCAGCAATTAAGCTGGCGGAA 2967
 1346 LysSerAsnGlyAlaThrAlaThrGluThrAspLysLysValAla... 1361
 2968 AGTTCGGAAGCACTTACACCTTGGCGGCTCAACATACCGGCAAC... 3012
 1362 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaIaThrPheVal 1377

3013 .GACCTGCAGACCTCGAACAATTGACGTAAGTGGAGAGAAAAGACACA 3061
 1377 LysValGluAsnAspSerAlaThrIle.....AspAsp 1390
 3062 AACCGCTGTCGGAACCTTAATTCACCCCTGCA...AACGACACGTC 3108
 1390 ePThrThrAspAspGlyAlaAsnAspAlaLeuLysAla***AspThrLeu 1406
 3109 GATCGCGCGCGCTGCGCTTACCAACTCATCCGCAACAGCGCGAG... 3153
 1407 ThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGlyLysAsnI 1423
 3154TTCCGCTGCAATTAAT.....CCGCTCAAGACACAGCTTCG 3193
 1423 eThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaThrValSerA 1440
 3194 ACAACTC.....GCCAAGCGCAAGAACCCAAACACG 3225
 1440 sPysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSerAsp 1456
 3226 GCGGAA.....AAGCAACAGCGCAAGGCTTGCAGCGCT 3260
 1457 ThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAlaAs 1473
 3261 GATTGCGCGCGCGCGCGATCCGCTGAAAGACAGAAACGCTT.....G 3304
 1473 nIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuLeuAsnG 1490
 3305 CCGAACCGCGCGCGCGACGCGGCAAAATGCGGATTTGTCAGCGC 3354
 1490 eGlyAlaThrThrAsnLeuGlyGlyAsnGlyIle.....Thr 1502
 3355 GAGGAAGAGAAAAAACGCGTGCAGCGGATTAAGACACCGCTTGCAGAA 3404
 1503 AspAsnGluLysLys***AlaAlaSerValLysAspValLeuAsnAlaGlu 1519
 3405 ACAGCGCGAAGCGGAACCCGCGGCTACACCGCTTCCCGCGCGCC 3454
 1519 yThrAsnValAlaArgGlyValLysPheAlaSer..... 1529
 3455 GCGCGCGCGCGCGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCA 3504
 1529 1529
 3505 CCGCAGCGGACCTGATACGCCGTTATGCAATAGCGGTTTGAGT.... 3549
 1530AlaAsnAsnGlnValGluAsnI 1537
 3550 .GAATTTCCGCGCAGCTCACACGCTTTTCCGCTACAGACGATTAAG 3598
 1537 eaSpPheValAlaThrTyrAspThrValAspPheValSerGlyAspLysA 1554
 3599 AC.....CGGTAATTTGCGGAAGACCGCGCAC..... 3627
 1554 sPThrThrSerValThrValGluSerLysAspAsnGlyLysArgThrGlu 1570
 3628GCCGTTTGACAAGCGCATCCGCGACACCAACACTACCG 3668
 1571 ValLysIleGlyAlaLysThrSerValIleLysAsp.....HisAs 1584
 3669 TTGCGCAAGATTTCCGCGCTACCGCCCAACAAACCACTCGCGCAAAACG 3718
 1584 nGlyLysLeuPheThrGly...LysGluLeuLysAspAlaAsnAsnAG 1600
 3719 GTATGACG.....AAAACCTGCGCGCGCGCGCTC 3750
 1600 LyValThrValThrGluThrAspGlyLysAspGluGlyAsnGlyLeuVal 1616
 3751GGCATCCTGTTTTCGCAACAACCG..... 3774
 1617 rThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTTrpArgVal 1633
 3774 3774

```

1633 srhThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrValAlaSerG 1650
3775 .ACCGAAACACCTTCGACGAGCGATCGGCACTCGCGAGCGCTTCC 3822
1650 LysThrAsnValThrPheAlaAspGlyAsnGlyThrThrAlaGluValThr 1666
3823 .....CAGCGCGCGCTTTCGGGCAATACGCGATCGACAGTTCCTA 3863
1667 LysAlaAsnAspGlySerIleThrValIstYThrVal..... 1679
3864 CATCGCGATCAGCGCGCGCGGCTTTTACGACGCGCGCTTCAGACG 3913
1680 .....LysValAlaAspG 1684
3914 GCATC.....GGAGCAAAATC 3930
1684 LysLeuLysLeuAspGlyAspLysIle 1692

```

```

seq_name: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:US-08-685-467-4
seq_documentation_block:
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robln M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

```

```

alignment_scores:
Quality: 218.50      Length: 1492
Ratio: 0.322         Gaps: 73
Percent Similarity: 45.442   Percent Identity: 19.035

```

```

alignment_block:
US-09-303-518D-649 x US-08-685-467-4

```

Align seg 1/1 to: US-08-685-467-4 from: 1 to: 1912

```

172 GCGCAAAATTAAGCAAGTTGTCAGTCGGGGCGAAAGTATTGAGTTA 221
435 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 450
222 CAACAAAAAGGGAGTGTGTC.....GGCAATCAATG...A 256
451 ...LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 466
257 CAAGACCCCGATGATGATTTCTGTGTGTGCGCGCAAGCGCGTGGCG 306
466 allYsGlnAspGlyAlaAsnPheThrIYrSerLeuGlnAspAlaLeuThr 482
307 GATGTGTGGGC..... 318
483 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 499
319 .....GATCAATATATTTGTGACGCTGCGACATAACG 349
499 alysthValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 516
350 GCGGCTATACACGCTGATTTGTGCGGAGAGAAATCCGATCAA 399
516 ly.....GlyThrThrGlyThrAsn..... 522
400 CATCGTTTACTTATAAATGTGAACGGAATATTTAAAGCAGGAGAC 449
523 .....ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 535
450 TAAAGCCATCCTTAWGGCGGATATCATATGCCGGTTCATTAAT 499
535 nlysaAla..... 537
500 TTGTACAGATCAGACCTGTTGAATGACCAGTTATATGATGGCGG 549
538 .....IleThrAsnValAlaSerIlyLeu 545
550 AAATATATGATCAAAATTAATTAACCTGACCGTGTGATTTGGGCGAG 599
546 ArgAlaTYrAspAspAlaAsnPhe..... 553
600 CAGGCAATATTTGGCATCTGATGAGATGAGCCCAATAAC.....CGCG 643
554 .....AspValLeuAsnAsnSerAlaThrA 562
644 AAAGTTCATATCATATTTGCAAGTGGGTATTTGGCTGCTGTGGCAAT 693
562 spleuAsnArghIysValGluAspAlaTYrLys..... 572
694 ACCTTTGCACAAATGATCAGTGTGTGCGACAGTCACTAGTAGTAGGA 743
573 .....GlyLeuLeuAsnLeuAsnGluLys 580
744 AAAAATTAACATAGCCCATATGCTTTTATACCAACAGAGGCTCATTTG 793
580 sasAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrValG 597
794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCAAAGCAAAAGTGG 843
597 ly.....AspLeuArgLysLeuGlyTrp 604
844 TTAATT.....AATGGGATTTGCAACGGCGCAACCTATATATAGG 884
605 ValValSerThrLysAsnGlyThrLysGluGluSerAsn..... 617
885 AAAAGCAATGGCTTCAGCTGTCGTAAGATTTGTTATGATGAGAA 934
618 .....GlnValLysGlnAlaAsp.....GluValL 626
935 TCCTTGCTGAGATACCATTCAGTATTTCTAGCAACGACGTCAAAATGGG 984

```

626 eupheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 642
985 AATACTCTTTTAAGCAGCATATATATGGCAGCAAAAATCAATGACCA 1034
643 LysHisThr..... 645
1035 ACATGACACAACTCTGCTAATAGATTAAACACGACCGTTCAAT 1084
645 645
1085 TGTATTATGTTCTTTATCCGACAGCAGCAAGCACTGTTATCATGCT 1134
646 ..LleThrValSerValAlaGlyThrLysAlaAspCysGlyLeuGluLys 661
1135 GCAGGTGGTGCACACGTTATCGACCGCAGACTGAATATGAGAAAATAT 1184
662 AspGlyAspThrLleLysLeuLysValAspAsnGlnAsnThrAspAsnVal 678
1185 TTCCTTATTT..... 1194
678 LleThrValGlyAsnAsnGlyThrAlaValThrLysGlyLypheGluT 695
1195GACGAGAAAAGCGAATTGATCT... 1221
695 hrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLys 711
1222ACCAGCAAC..... 1230
712 AspAlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspVal 728
1231ATCAATCAAGTGCTGCGAGATTATATTTCCAGAGAGATT 1270
728 LAlaThrAlaLleAsnSerAlaAlaThrPheValLysThrGluAsnLeuT 745
1271 TTACGCTCCGCTGAAATATACGAACTGGCAAGGCGGCGGTTTCAT 1320
745 hrThrSerLleAspLysAspAsnProThrAspAsnGlyLysAspAspAla 761
1321 ATCACTGAGACAGTACCGTTACTTGAAAGTAAACGCGCGCAACGA 1370
762 LeuLysAlaGlyAspThrLeuThrPheLysAlaGly.....LysAsnLe 776
1371 CCGCGTGCACAAATCGGCAAGGC..... 1395
776 uLysValLysAspGlyLysAsnLleThrPheAspLeuAlaLysAsnL 793
1396ACGCTGACGTTCAAGCCAAA 1416
793 euGluValLysThrAlaLysValSerAspThrLeuThrLleGlyLys 809
1417 GGGGAAAACCAAGGCTG.....ATCAGCGT 1442
810 ThrProThrGlyGlyThrThrAlaThrProLysValAsnLleThrSerTh 826
1443 GGGGAGCGTACAGTCATTTGGATCGACGCGCAGCATTAAGCCAAA 1492
826 rAlaAspLleuAsnPheAlaLysGluThrAlaAspAlaSerLysThr 843
1493 AACAGGCTTT.....AGTGAATC 1512
843 yAsnValLysThrLysGlyLleAlaThrThrLeuThrGluProSerAla 859
1513 GCGTTGGTACGCGGAGGCGTACGCGCAACTGAATCCGATTAATCAGTT 1562
860 GlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysLys 876
1563 CAACCCGCAAACTATTCGGCTTCGGGGCGGA...CGTTGAT 1609
876 rAsnAlaLaserLleGluAspValLleuArgAlaGlyTTPAsnLleGln 893
1610 TAAACGGGCGATTCGTTCCGTCACCGTATTCAAAATCCGATGAAGG 1659
893 LysAsnGlyAsnAsnValAspLysValAlaThrThrLysThr..... 906
1660 GCGATGATTTGTCACCAACAATCAAGCAAAAGATCCAGCGTTACATTAC 1709
907ValAsnPheThrAspSerThrGlyThrThrValThr 920
1710 AGCGAATTAAGATATTGCTAACCCGCAATATACACAGCTTGATAGCA 1759
920 rValThrGlnLysAlaAspLysGlyAlaAspValLysLleGlyAlaL 937
1760 AAAAGCAATTGCC.....TACACGCT...TGGTTGGCAGAAAAGT 1800
937 yThrSerValLleLysAspHisAsnGlyLysLeuPheThrGlyLysAsp 953
1801 ACAGCAAAAAGCAAGGGGGGCGTCAACCTGTTTACCAGCCCGCGGCA 1850
954 LeuLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspLysLys 970
1851 AGACCGCACCGCTGCTCTTCGGCGGACA.....AAT 1885
970 rThrGlyThrGlyLeuValThrAlaLysThrValLleAspAlaValAsnL 987
1886 TAAACGGCAACATCACGCCAAACAGCAAACTGTTTTCAGCGGCA 1935
987 ySerGlyTTPArgValThrGlyGluGlyAlaThrAlaGlyThrGlyAla 1003
1936 CCACACCGCAGCGCTACATCAATTAACGACCAATGCTGCCAAAAGA 1985
1004 ThrAlaValAsnAlaGlyAsn.....AlaG 1012
1986 GGGCATTCCTCGGGGGAATCGTGTGGACAGACGTGATCAACCGCA 2035
1012 uThrValThrSerGlyThrSerVal..... 1020
2036 CATTTAAACGGGAAACTCCAAATTAAGCGGAGCAGCGGTGCTTCC 2085
1021AsnPhe.....LysAsnGlyAsnAlaThrThrAla 1030
2086 CGCATGTTGCCAAAGTGAAGCGCATTTGGCATTTGAGCAATCACGCCCA 2135
1031 ...ThrValSerLysAspAsnGlyAsnLleAsnValLysTyrAspValAs 1046
2136 AGCAGTTTGGTGTGCGCACCGCATCAAGCCACACATCTGTACAGCT 2185
1046 nValGlyAspGlyLeuLysLleGlyAspAspLysLleVal.....A 1061
2186 CGGACTGAGCGGCTGACCAAAATGTGTGCAAAAACATTCACGACGAT 2235
1061 laAspThrThrThrLeu.....ThrValThrGlyGly 1071
2236 AAGTGTATGCTTCATTCAGTAAAGCCGACATCAGCGCATGTGATCT 2285
1072 LysValSerValProAlaGlyAlaAsnSerValAsnAsnLysLysLe 1088
2286 TGGCGATCAGCTCATTTAATCTCACAGGCTTCCACACTCAACGCA 2335
1088 u.....ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1100
2336 ATCTTAGT.....GCAATGGGATACAGT 2361
1100 snLeuSerThrPheAlaLysAlaAspLysTyrAlaAspLysValuSerGlu 1116
2362 TATACAGTCAGCCACAGCCACCCAAAAGGCAACTAGCCTC... 2406
1117 GlyLysThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAl 1133
2407GTGGCAATGCCCAAGCAACATTTAAT... 2433
1133 eGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1150
2434 ...CAAGCCCATTTAAAGCGCAACATCGGCTG...GGCATGCT 2475
1150 euGlnAspThrLeuThrGlyLeuThrSerLleThrLeuGlyGlyThrAla 1166

```

2476 TCATTTAATCTAAGCGACCGCCGTACAAAGCGC.....AGCTCT 2516
      ::::: ::::: |||
1167 AsnGlyAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1183
      ::::: ::::: |||
2517 GACGCTTCCGGCAACGCTAAGCAACGTAAGCATTCGGCTACG 2566
      ::::: ::::: |||
1183 eThrLeuLysAsnGlyAlaAlaIleGlyThr.....AspAlaSerAsn 1198
      ::::: ::::: |||
2567 GTATT.....GTCCTCAGCGCATAGCA 2592
      |||
1198 LysAsnThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1214
      ::::: ::::: |||
2593 GTATTCATTTTGAAGCAGCGCTTACCGGACCAATCAGCGCGGCA 2642
      ::::: ::::: |||
1215 IleThrAsnValLysSerAlaLeuLysThr.....TyrLysAspThrG 1229
      ::::: ::::: |||
2643 GGATACGGCATTTACATTAAAGACACGAAATG.....ACGTCGGCT 2686
      ::::: ::::: |||
1229 AsnThrLysAspGlyThrGlnAspLysGluPheHisAlaValLysA 1246
      ::::: ::::: |||
2687 CAGCAGCAGAAATTAGCAATTTAACTTGACACGCGCACCATTA 2733
      ::::: ::::: |||
1246 snAlaAsnGlyValGluPheValGlyLysAsnGlyAlaThrValSerAla 1262
      ::::: ::::: |||
2734 ..CTCAATTCGGCTATCGCCAC.....GATCGCGCAGGGGC 2768
      ::::: ::::: |||
1263 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1279
      ::::: ::::: |||
2769 GCAACCGCGCGTGGC.....ACAGATGCG.....CCGCGCC 2800
      ::::: ::::: |||
1279 aLysValGlyAspGlyLeuGlyLysAspThrAspGlyLysIleLysLeu 1296
      ::::: ::::: |||
2801 GCGCTGCGCGCTTCCGCGCTCCCTATTATCCGTT.....ACACCG 2844
      ::::: ::::: |||
1296 ySValAspAsnThrAspGlyAsnAsnLeuThrValAspAlaThrLys 1312
      ::::: ::::: |||
2845 CCAACTTCGCTA...GAATCCCGTTTCAACACGCTGAC..... 2880
      ::::: ::::: |||
1313 GlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaTh 1329
      ::::: ::::: |||
2881 .....GTAAACGCAATTGAACGCTACGCGCAACATTCGCT 2917
      ::::: ::::: |||
1329 rThAlaGlnGlyThrAsnAlaAsnGluAlaGlyLysValValLysG 1346
      ::::: ::::: |||
2918 TTATGTCGAATCTTCGCGACGACGACGCAATGATGACGCGGAA 2967
      ::::: ::::: |||
1346 LysSerAsnGlyAlaThrAlaThrGluThrAspLysLysValAla... 1361
      ::::: ::::: |||
2968 AGTTCGAGAGCACTTACACCTTGGCGGTCAACATACGCGCAAC... 3012
      ::::: ::::: |||
1362 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaIleThrPheVa 1377
      ::::: ::::: |||
3013 .GAACCTGCAAGCCTCGACACATGACGTAAGTGAAGAAAGACACACA 3061
      ::::: ::::: |||
1377 LysValGluLysAsnAspAspSerAlaThrIle.....AspAsp 1390
      ::::: ::::: |||
3062 AACCGTGTGCGAAACCTTAATTTCACCTGCA...AACGAACACGTC 3108
      ::::: ::::: |||
1390 ePThrAspAspGlyAlaAsnAspAlaLeuLysAla***AspThrLeu 1406
      ::::: ::::: |||
3109 GATCCGCGCGGTGGCTTACCACTCATCCGCAAGACGCGGAG... 3153
      ::::: ::::: |||
1407 ThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGlyLysAsnI 1423
      ::::: ::::: |||
3154 ...TTCGGCTGCATTAAT.....CGGTCAAGACACAGAGCTTCCG 3193
      ::::: ::::: |||
1423 eThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaThrValSerA 1440
      ::::: ::::: |||
3194 ACAAACTC.....GGCAAGCGACAAGCAAAACAAACAG 3225
      ::::: ::::: |||
1440 sPlysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSerAsp 1456
      ::::: ::::: |||
3226 GCGGA.....AAGACACACGCGCAAGCCTTGACGCGCT 3260

```

```

1457 ThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAlaAs 1473
      ::::: ::::: |||
3261 GATTCGCGCGCGGCGCATGCCGTGCAAAAGACAGAAAGCTT.....G 3304
      ::::: ::::: |||
1473 nIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuLeuAsn 1490
      ::::: ::::: |||
3305 CCGAACCGCGCGCGGACGCGGGAATGTGGCATTTATGACAGCG 3354
      ::::: ::::: |||
1490 eGlyAlaThrThrAsnLeuGlyGlyAsnGlyIle.....Thr 1502
      ::::: ::::: |||
3355 GAGGAAGAGAAAAACGCGTGCAGCGGATTAAGACACCGCTTGCGGA 3404
      ::::: ::::: |||
1503 AspAsnGluLysLys***AlaAlaSerValLysAspValLeuAsnAlaG 1519
      ::::: ::::: |||
3405 ACAGCCCGAAGCGGAACCGCGCGCTACACCGCTTCCCGCGGCC 3454
      ::::: ::::: |||
1519 yTPAsnValArgGlyValLysProAlaSer..... 1529
      ::::: ::::: |||
3455 GCGCGCGCGCGGATTTGCGCAACTGCAACCCCAACCGCACGCCCA 3504
      ::::: ::::: |||
1529 ..... 1529
      ::::: ::::: |||
3505 CCGCAGCGGACCTGATCAGCCGTTATGCCAATAGCGGTTGAGT..... 3549
      ::::: ::::: |||
1530 .....AlaAsnAsnGluValGluAsnI 1537
      ::::: ::::: |||
3550 .GAATTTCCGCGCAGCTCAACAGCGTTTCCCGCTACAGACGAATTAG 3598
      ::::: ::::: |||
1537 eAspPheValAlaThrTyrAspThrValAspPheValSerGlyAspLysA 1554
      ::::: ::::: |||
3599 AC.....CGGTAATTGCGGAAGACCGCGGCAC..... 3627
      ::::: ::::: |||
1554 sPThrThrSerValThrValGluSerLysAspAsnGlyLysArgThrGlu 1570
      ::::: ::::: |||
3628 .....GCCGTTTGACAAAGCGCATCGCGGACACCAACACTACCG 3668
      ::::: ::::: |||
1571 ValLysIleGlyAlaLysThrSerValIleLysAsp.....HisAs 1584
      ::::: ::::: |||
3669 TTCGCAAGATTTCGCGCGCTACCGCCACCAACCAACCTCGGCCAAATCG 3718
      ::::: ::::: |||
1584 nGlyLysLeuPheThrGly...LysGluLeuLysAspAlaAsnAsnAsn 1600
      ::::: ::::: |||
3719 GTATGCAG.....AAAACCTGCGGACGCGGCGCGC 3750
      ::::: ::::: |||
1600 LysAlaThrValThrGluThrAspGlyLysAspGluGlyAsnGlyLeuVal 1616
      ::::: ::::: |||
3751 .....GGCATCCTGTTTCGCACAACCG..... 3774
      ::::: ::::: |||
1617 ThrAlaLysAlaValIleAspAlaValaLeuLysAlaGlyTPArgValLy 1633
      ::::: ::::: |||
3774 ..... 3774
      ::::: ::::: |||
1633 sThrThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrValAlaSerG 1650
      ::::: ::::: |||
3775 .ACGGAACACACTTCAGACGACGATCGCACTCGCAGCGCTTCCG 3822
      ::::: ::::: |||
1650 LysThrAsnValThrPheAlaAspGlyAsnGlyThrThrAlaGluValThr 1666
      ::::: ::::: |||
3823 .....CACGCGCGCTTTCGGGCATACGCGCATCGACAGGTTCTA 3863
      ::::: ::::: |||
1667 LysAlaAsnAspGlySerIleThrValLysTyrAsnVal..... 1679
      ::::: ::::: |||
3864 CATCGGCATCAGCGCGGCGGTTTATGACGCGGACGCTTTCAGAG 3913
      ::::: ::::: |||
1680 .....LysValAlaAspG 1684
      ::::: ::::: |||
3914 GCATC.....GGAGCAAAATC 3930
      ::::: ::::: |||
1684 LysLeuLysLeuAspGlyAspLysIle 1692

```

seq_name: /cgn2_6/ptodata/1/1ae/5A_COMB.pep:US-08-409-995-2

```

seq_documentation_block:
: Sequence 2, Application US/08409995
: Patent No. 5646259
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen I.
: APPLICANT: St. Geme III, Joseph W.
: TITLE OF INVENTION: Haemophilus Adhesion Proteins
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Test, Albritten & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/409,995
: FILING DATE: 24-MAR-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-61053/RFT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1098 amino acids
: TYPE: amino acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: US-08-409-995-2

alignment_scores:
: Quality: 214.00 Length: 1396
: Ratio: 0.347 Gaps: 72
Percent Similarity: 44.198 Percent Identity: 20.129

alignment_block:
US-09-303-518D-649 x US-08-409-995-2 ..

Align seg 1/1 to: US-08-409-995-2 from: 1 to: 1098

22 ACAACGAAACACACGCGAAAGCCGCGCATCCGCTTC 71
|||||:|||||:|||||:|||||:|||||:|||||:
46 ThrValGluAlaAsnAsnAsnThrProValThrAsnLysLeuLys.... 60
72 GCCTGCTTACTTAGCCATATGCTGCTGCTGCGCATTCCTCCCAAGCCT 121
|||||
61 ...AlaTyr..... 62
122 GCGGCGGACACACTATTTCGGCATCACTACCAATCTATCGGCATT 171
|||||:|||||:|||||:|||||:|||||:|||||
63 ....GlyAspAlaAsnPhenAsnPhenThrAsnAsnSerIleAlaAspAla 77
172 GCCGAATATAAGGCAAGTTTGCAGTCGCGGCGCAAGATATTGAGGTTA 221
|||||:|||||:|||||:|||||:|||||:|||||
78 GluLysGlnValGlnGlnAlaTyrLysGlyLeuLeuAsnLeuAsnGlnLys 94
222 CAACAAAAAAGGAGTGTGCGCAATCATGACAAAAAGCCCGCATG. 270
|||||:|||||:|||||:|||||:|||||:|||||
94 sAsnAlaSerAspLysLeuLeuValGluAspAsnThrAlaAlaThrValG 111
271 .....ATTGATTTTCTGTGTGTCGCTGACGCGCGGCGGCG 306
|||||:|||||:|||||:|||||:|||||:|||||

```

```

111 LysAsnLeuArgLysLeuGlyTyrPValLeuSerSerLysAsnGlyThrArg 127
307 GCATTGTGGCGGATCATATATTTGTAGCGGACAT..... 345
128 AsnGlnLysSerGlnGln.....ValLysHisAlaAspGlnVal 140
346 .....AACGGCGGCTATACACACGTTGATTTGGTCGCGAAG 382
140 LleupheGlnGlyLysGlyGlyValGlnValThrSerThrSerGlnAsn 157
383 GAAGAAATCCCGATCAACATGCTTTACTTATTAATTTGTAAACGAGAT 432
157 Lys.....HisThrIleThrPheAlaLeuAlaLysAspLeu 169
433 AATTATTAAGCAGGAGCT.....AAGGCGCATCC 461
170 GlyValLysThrAlaThrValSerAspThrLeuThrIleGlyGlyAla 186
462 TTATGCGCGCATTTATCATATGCGCGCTTGCAATTAATTTGTACAGATG 511
186 AlaAlaLys.....AlaThrThrT 193
512 CAGAACCTGTGAATGACAGTATATGATGAGCGGGAATATATGAT 561
193 hrProLysValAsnValThrSerThrThrAspGlyLeuLysPhe..... 207
562 CAAATTAATTACCTGACCTGACCTGCTGATTTGGCGGCGACGACATATG 611
208 .....AlaLysAspAlaAlaGlyAla..... 214
612 GCGATCTGATAGATGAGCCCAATACCGGCAAGTTCAATCATATG 661
215 .....AsnGlyAspThrThrValHisLeu. 222
662 CAAGTGGCATTTCTTGCGCTGTGTGGCAATACCTTGACAAAATGA 711
223 .....AsnGly 224
712 TCAGCTGTGCGCACATCACTTAGTAGTAGAAAAAATTAACATAGCC 761
225 IleGlySerThrLeu.....ThrAspThrLeuValGlySerPr 237
762 ATATGCTTTTACCACACGAGGCTCATTTGGCGACAGTGGCTACCAA 811
237 AlaThrHisIleAspGlyLysAspGlnSerThrHisTyrThrArgAla 254
812 TGTATTATCATGATGCCCAAAAGCAAAAGTATTAATGAGGATATG 861
254 LAserIleLysAspValLeuAsnAlaGlyTTPAsnIleLysGlyAla... 269
862 CAACGGGCGACCCCTATATAGGAAAAAGCAATGGCTTCACAGTGGTCG 911
270 LysAlaGlySerThr...ThrGlyLysSerGlnAsnGlyLysArgThr. 285
912 TAAAGATTGTTCTATGAT.....GAATCTTTGCTGGAATACCAT. 954
285 sThr.....TyrAspThrValGlnPheLeuSerIleAspThrGlnT 299
955 .....TCAGTATTTCAGACACAGCTCAAAATGGCAATFACCTTTTAC 999
299 hrThrThrValThrValAspSerLysGlnAsnGlyLysArgThr..... 313
1000 GACGATTAATATGCGACAGAAAAATCATGCAAAACATGAACACAAATTC 1049
314 .....GluValLysIleGlyLys..... 320
1050 TCTGCCATATAGATTAAACACAGACCGTTCAATGTTTATGTTTCTT 1099
321 ....ThrSerValIleLysGlnLysAspGlyLysLeuPhe..... 332
1100 TTATCGAGACAGACAGACAGACCTGTTTATCATGCTGCGAGTGGTGTAC 1149
333 .....ThrGlyLysAlaAsn 337

```



```
1150 AGTTATCGACCCAGCTGATAATGAGAAATATTCTTATTGACGA 1199
    ::::::::::::::::::::: |||
338 LysGluThrAsnLysValAspGlyAlaAsnAlaThrGluAspAlaAsp 354
1200 AGGAAAGCGCAATTGATCTACAGCAATC..... 1233
    |||
354 uGlyLysGly...LeuValThrAlaLysAspValIleAspAlaValAsnL 370
1234 .....AATCAAGTCGCTGGAGATTATTTCCAA 1263
    ::|
370 ySThrGlyTPArgIleLysThrThrAspAlaAsnGlyLysLysLysLys 384
    ::|
1264 GGAGATTATTT...ACGGTCGCGCTGCAAAATTAAGCAAACTTGG...CAAG 1307
    |||
385 GlyAspPheAlaThrValAlaSerGlyThrAsnValThrPheAlaSerG 401
1308 CGCGGCGCTCATATCATGTAAGACAGTACCGCTTACTTGAAAGTAACG 1357
    |||
401 yAsnGlyThr.....ThrAlaThrValThrAsnGlyThrAspG 414
1358 GCGTGGCAACGACCGCTGTCCAAATGCGCAAGCGACGCTGCACGTT 1407
    |||
414 LylleThrValLysTyrAspAlaLysValGlyAspGly.....Leu 427
    ::|
1408 CAAGCCAAAGGGGAA.....AACCAAGCTCGTCAGCTGCG 1445
    ::|
428 LysLeuAspLysAspLysIleAlaAlaAspThrThrAlaLeuThrValAs 444
1446 CGACGCTACGTCATTTGGATGAGCAGCAGCAGATTAAGGCAAA... 1491
    |||
444 naSpGly.....LysAsnAlaAsnAsnProLysGlyLysValAla 457
1492 ....AACAAGCCTTAGCAATGCGCTGTCAGCGCGCAGGCGGTACG 1536
    |||
457 laAspValAlaSerThrAspGluLysLysLeuValThrAlaLysGlyLeu 473
1537 GTGCACTGATGATCCGATATCAGTTCAACCCGACAACTCATTTCCG 1586
    |||
474 Val.....ThrAlaLeuAsnSerLeuSerTrpThrThrAlaAlaG 488
1587 CTTTCGCGCGGACGTTTGATTTAAAC..... 1614
    |||
488 uAlaAspGlyLysThrLeuAspGlyAsnAlaSerGluGluValLysAla 505
1615 ..GGGCACTTGGCTTGTCCACCGATTCAAAT.....ACCGAT 1653
    |||
505 laGlyAspLysValThrPheLysAlaGlyLysAsnLeuLysValLysGlu 521
1654 GAAGGCGCGATGATGTCACACCAATCAAGCAAGAATCCACGCTTAC 1703
    |||
522 GluGlyAlaAsnPheThrLysSerLeuGluAsp.....Al 533
1704 CATTACAGCAATTAAGATATGCTACA...ACCGCAATTAACACAGCT 1750
    |||
533 aLeuThrGlyLeuThrSerIleThrLeuGlyThrGlyAsnAsnGly.... 548
1751 TGGATACCAAAAAAGAAATTCCTACACGCTTGCTTGGCGGAAAGAT 1800
    |||
549 ....AlaLysThrGluIleAsnLysAspGly.....LeuThr 559
1801 ACGACCAAAAGCAAGCGCGCTCAACCTTGTTTACACCGCCCGCAG 1850
    |||
560 IleThrProAlaAsnGly..... 565
1851 AGACCGCACCCCTGCTGCTTCGCGCGCAACAATTAACGCAACATCA 1900
    |||
566 .....AlaGlyAlaAsnAsnAlaAsn...ThrIle 575
1901 CGCAAAACAAACGCAAACTGTTTTCAGCGCGCAGACCAACGCGCAGCC 1950
    ::|
575 eValThr..... 577
```

```
1951 TACATCATTTAAACGACCATTTGTCGCAAAAAGAGGCGCATTCCTCGCG 2000
    |||
578 .....LysAspGlyIleSer..... 582
2001 GGAAATCGTGGGACAACGACTGATCACCAGCATTTAAAGCGGAA 2050
582 ..... 582
2051 ACTTCCAAATTAAGCGGACAGCGCGTTCCTCCGCAATGTTGCCAA 2100
    |||
583 .....AlaGlyGlyIleSerVal.....LysAsnVal..... 591
2101 GTAAAGCGCATTTGGCATTTGAGCAATCAGCCCAAGCATTTTGTGCT 2150
    |||
592 ValSerGly.....LeuLysLysPheGlyAspAlaAsnPheAspPr 605
2151 GCGACCGCATCAAGCCACACAAATCTGTACAGCTTGGAC.....TGA 2194
    |||
605 OLeuThrSerSerAlaAspAsnLeuThrLysGluAsnAspAspAlaTyrL 622
2195 CGGCTCGCAAAATGTGTCGAAAA...ACCATTCAGCAGATTAAGTG 2241
    |||
622 ySgLYLeuThrAsnLeuAspGlyLysGlyThrAspLysGluThrProVal 638
2242 ATTGCTCATTTGACTAAGACGACATCAGCGCAATGTGATCTTGCGCA 2291
    |||
639 ValAlaAspAsnThrAlaAlaThrValGly..... 648
2292 TCACGCTCATTTAAATCTCACAGGCTTGGC...ACACTCAAGCGCAATC 2338
    |||
649 .....AspLeuArgGlyLeuGlyTyrValIleSerAlaAspL 661
2339 TTAGTGCAAAATGCGATACAGTTATACAGTCAGCAGCAACGCCCA 2388
    |||
661 ySThrThrGlyLysSerThrGluTyr.....HisAspGluValArg 674
2389 AACGGCAACCTTAGCCCTGCGGCAATGCCAGCAACATTTAATCAGC 2438
    |||
675 AsnAlaAsn.....GluValLysPheLys..... 682
2439 CACATTTAAAGGCAACACATGCGCTTGGCGCAATGCTTATTAATCA 2488
    |||
683 .....SerGlyAsn...GlyIleAsnValS 690
2489 GCGACACGCGCGTACAAAAGCGCAGTGTGACGCTTCCGCAACGCTAG 2538
    |||
690 eGlyLysThrVal.....AsnGlyArg 697
2539 GCAACGTAAGCCATTCGACATCAACGTAATGTCTCCAGCGCATTA 2588
    |||
698 ArgGluIleThrPheGluLeuAlaLysGlyGluVal..... 709
2589 GGCAGTATTCATTTGAAGCAGCGCTTACGCGCAAAATCAAGCGCG 2638
    |||
710 .....ValLysSerAsnGluPheThrValLysGluThrAsnG 722
2639 GCAAGGATACGGCATTA..... 2655
    |||
722 LysGluThrSerLeuValLysValGlyAspLysTyrTyrSerLysGlu 738
2656 .....CACTTAAGAAGACAGCATG... 2676
739 AspIleAspLeuThrThrGlyLysProLysLeuLysAspGlyAsnThrVa 755
2677 .....ACGCTGCCGTGAG 2689
755 lAlaAlaLysTyrGluAspLysGlyLysValValSerValThrAspA 772
2690 GCAAGCAATTAAGCAATTAACCTTGACAAAGCCACCATTAACATCA 2739
    |||
772 snThrGluAlaThrIleThrAsnLysGlySerGlyTyrValThrLysLys 788
2740 TCCGCTATCGCCACGATGCG.....GCAGGGGCGCAAAACGCGCAG 2780
```

